

SKEW T BASED MAXIMUM LIKELIHOOD ESTIMATION OF LATENT GROWTH
CURVE MODELS WITH NON-NORMAL AND MISSING DATA

A Dissertation

by

MARIA ANTOUN HENRI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Oi-Man Kwok, PhD
Committee Members,	Myeongsun Yoon, PhD
	Wen Luo, PhD
	Lei-Shih Chen, PhD
Head of Department,	Shanna Hagan-Burke, PhD

December 2018

Major Subject: Educational Psychology

Copyright 2018 Maria Henri

ABSTRACT

Latent growth curve models (LGM) are widely used in educational research to analyze longitudinal data. Typical normal-based maximum likelihood estimation (n MLE) assumes that data are normally distributed. Violations to the normality assumption have grave consequences on the accuracy of parameter estimates, which are augmented when missing data are present. Several robust modifications have been proposed to remedy the effects of the violation of the normality assumptions, the most common being robust normal based maximum likelihood (n MLR). However, these methods have serious limitations. Assuming that the data follow skew t distribution within the maximum likelihood framework (st MLE) provides a more parsimonious alternative. Recently, Mplus has implemented a distribution option that makes implementing st MLE more feasible.

This study was conducted to evaluate the performance of st MLE in the estimation of LGM through a Monte Carlo simulation. Application of st MLE was also illustrated through estimation of LGM with math achievement test data from the National Longitudinal Survey of Youth. Results confirmed that n MLR can still produce biased parameter estimates when data are non-normally distributed. On the other hand, st MLE resulted in many estimation issues. Although st MLE presents a theoretically appropriate framework to estimate LGM with non-normal data, more research is needed to determine the conditions under which it performs well.

DEDICATION

To Dillon McCameron for always giving me hope, even when my models wouldn't converge.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Oi-Man Kwok, for being a great mentor to me throughout my time as a student. I am very grateful for his dedication to helping his students reach their goals. I would also like to thank my committee members, Dr. Luo, Dr. Yoon, and Dr. Chen for their guidance and support. I am very thankful for all Research, Measurement, and Statistics faculty members and colleagues, I learned so much from them about being a good researcher. I would like to thank my RMS colleagues, especially Brandie Semma, Lucy Chen, and Mirim Kim; their friendship filled my grad school experience with laughter, good food, and unforgettable memories.

CONTRIBUTORS AND FUDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Oi-Man Kwok and Professors Myeongsun Yoon and Wen Luo of the Department of Educational Psychology and Professor Lei-Shih Chen of the Department of Health and Kinesiology

All work for the dissertation was completed independently by the student.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
CONTRIBUTORS AND FUDING SOURCES	v
TABLE OF CONTENTS	vi
LIST OF FIGURES.....	viii
LIST OF TABLES	ix
1. INTRODUCTION.....	1
Latent Growth Curve Models.....	3
Maximum Likelihood Estimation	8
Full Information Maximum Likelihood and Missing Data	11
Effects of Non-normality on n MLE Results	11
Robust Modifications	13
Multivariate Skew t Distribution.....	16
Examination and Application of st MLE	19
Gaps in the Literature and Purpose of the Study	20
2. METHOD: EMPIRICAL EXAMPLE	22
Participants and Data Structure	22
Model Specification and Analysis.....	23
3. RESULTS: EMPIRICAL EXAMPLE.....	25
4. METHOD: MONTE CARLO SIMULATION	27
Population Growth Model	27
Simulation Conditions	27
Evaluation Criteria	30

5. RESULTS: MONTE CARLO SIMULATION	32
Convergence Percentages.....	32
Parameter Relative Bias	34
Mean Square Error of Parameter Estimates	39
Coverage: 95% Confidence Interval	42
Type I Error	47
6. DISCUSSION	49
Summary of Results	49
Significance and Implications	50
Limitations and Future Research.....	53
7. CONCLUSION	55
REFERENCES.....	56
APPENDIX	67

LIST OF FIGURES

	Page
Figure 1 Confirmatory Factor Analysis (CFA) Model. A two-factor model with three indicators per factor and no cross loadings.....	80
Figure 2 Latent Growth Curve Model (LGM). Linear growth and uncorrelated residuals and six measurement occasions	80
Figure 3 Histograms of observed PIAT scores at each time point.....	81
Figure 4 Individual observed growth plots of PIAT across five time points for 200 randomly selected individuals	81
Figure 5 Random effects relative bias for growth conditions by estimation method.....	82
Figure 6 Growth factors covariance relative bias for zero growth conditions by estimation method.....	83
Figure 7 Fixed effects relative bias for growth conditions by estimation method	84
Figure 8 Slope mean relative bias for growth conditions by estimation method	85
Figure 9 Difference between estimated skewness and population skewness for Growth Condition..	86
Figure 10 Random effects MSE for growth conditions by estimation method.....	87
Figure 11 Growth factors covariance MSE for zero growth conditions by estimation method..	88
Figure 12 Fixed effects MSE for growth conditions by estimation method	89
Figure 13 Slope mean MSE for zero growth conditions by estimation method	90
Figure 14 Random effects power rates for growth conditions by estimation method	91
Figure 15 Fixed effects power for growth conditions by estimation method	92

LIST OF TABLES

	Page
Table 1 Empirical Illustration - PIAT Descriptive Statistics	67
Table 2 PIAT Tests of Normality.....	67
Table 3 Covariance Coverage	68
Table 4 Empirical Illustration Results - <i>st</i> MLE and <i>n</i> MLR Comparison	69
Table 5 Simulation Conditions.....	70
Table 6 Convergence rates for <i>st</i> MLE across all Conditions	71
Table 7 Intercept Variance Coverage Across all Growth Conditions	72
Table 8 Slope Variance Coverage Across all Growth Conditions	73
Table 9 Intercept/Slope Covariance Coverage Across all Growth Conditions	74
Table 10 Mean Residual Coverage Across all Growth Conditions.....	75
Table 11 Intercept Mean Coverage Across all Growth Conditions	76
Table 12 Slope Mean Coverage Across all Growth Conditions	77
Table 13 Intercept/Slope Covariance Type I Error Rates Across all Zero Growth Conditions.....	78
Table 14 Slope Mean Type I Error Rates Across all Zero Growth Condition.....	79

1. INTRODUCTION

The collection and analysis of longitudinal data is very common in educational research. Longitudinal data analysis is often used to estimate average shape and rate of growth trajectories, as well as examine growth processes and psychometric properties of measures over time. Latent growth curve modeling (LGM) is one of the most common methods of analyzing longitudinal data in educational research (e.g. Farley, Anderson, Irvin, & Tindal, 2017; Frischkorn, Greiff, & Wustenberg, 2014; Gottfried, Marcoulides, Gottfried, & Oliver, 2009; Mok & McInerney, 2015; Petersen & Hyde, 2017; Phan, 2012; You & Sharkey, 2009).

LGM is a flexible structural equation modeling (SEM) approach that allows the intercept and slope(s) of growth on a measure to be estimated as latent variables. LGM is commonly estimated using normal based maximum likelihood estimation (*nMLE*), which is built in many commercially available software, including SAS PROC MIXED, Mplus, and several R packages (e.g. lavaan; Rosseel, 2012). *nMLE* assumes that the endogenous and latent variables follow multivariate normal distributions (Hancock & Mueller, 2013; Kline, 2015; Preacher, Wichman, MacCallum, & Briggs, 2008).

Despite its popularity, *nMLE* makes assumptions that longitudinal data in the education field often violates. It has been well established that non-normality of real data is the rule rather than the exception (e.g. Cain, Zhang, & Yuan, 2017; Miccéri, 1989). Non-normality of observed longitudinal data indicates that either the growth latent variables and/or the residual terms are non-normally distributed (Brandt & Klein, 2015; Tong & Zhang, 2012). Violations of the assumptions of normality while using *nMLE*,

especially when coupled with missing data, may yield biased parameter estimates and standard errors, and unreliable model fit indices (e.g. Bollen, 2009; Shin, Davidson, & Long, 2009).

Over the past few decades, several methods have been proposed to make the estimation of SEM models robust to violations of the normality assumption. Each approach has its limitation, but what they all share in common is that they reduce the data to means and variances/covariances while disregarding higher order information about the shape of the data. Additionally, most robust methods rely on the assumption that the data are complete. Complete data is usually an unattainable goal, especially with longitudinal research, due to attrition (Li & Lomax, 2017). Furthermore, most of these methods make modification to the standard errors and model fit only, operating under the assumption that non-normality does not bias parameter estimates.

Recently, the application of skew t distribution within the MLE framework in the estimation of SEM based models has been receiving more attention. Skew t MLE ($stMLE$) has recently been incorporated as an option in the commercial statistical software Mplus (Mplus version 7.2; Muthén & Muthén, 2016). Research on the performance of $stMLE$ in the estimation of LGM with non-normal and missing data is limited, and the option in Mplus remains experimental (Muthén & Muthén, 2016). There is evidence to suggest that the $stMLE$ may outperform other robust modifications, under certain conditions.

In this study, the performance of $stMLE$ in the estimation of LGM with non-normal and missing data was examined through a Monte Carlo simulation. Specifically,

the performance of *st*MLE was compared to the most common robust method, normal based robust maximum likelihood (*n*MLR), under different distributional conditions.

The *st*MLE estimation of LGM with the Peabody Individual Achievement Test Mathematics scores from the National Longitudinal Survey of Youth, 1997 Cohort, was carried out to demonstrate the application of *st*MLE to real data.

Latent Growth Curve Models

LGM is a special case of SEM that can estimate shape and rate of growth of a construct over multiple measurement occasions. Proposed by Meredith and Tisak (1990), LGM is essentially specified as a confirmatory factor analysis (CFA) model, where factor loadings are fixed to reflect the hypothesized pattern of growth. The means and variances of the growth latent variables serve as estimates of average and variability of growth trajectories. To demonstrate how LGM is a special case of CFA, an example of a two-factor model is presented. A standardized two factor CFA, with three indicators per factor, no cross loadings, and independent heteroscedastic residuals (Figure 1) is represented in matrix form as:

$$X = \tau + \Lambda \xi + \delta$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

where τ is a vector of intercepts for each endogenous variable, X is a vector of endogenous variables, λ is a matrix of factor loadings, ξ is a vector of latent factors, and

lastly δ is a vector of residuals. The model implied variance/covariance matrix is represented as:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta$$

$$\Phi = \begin{bmatrix} \varphi_{11} & \varphi_{12} \\ \varphi_{21} & \varphi_{22} \end{bmatrix} \quad \Theta = \begin{bmatrix} \theta_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \theta_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} \end{bmatrix}$$

where Λ is a vector of factor loadings, Φ is the latent factors' variance/covariance matrix, and Θ is the residuals' variance/covariance matrix. The same CFA model illustrated above can also be shown in linear form as:

$$X_{ij} = \tau_j + \lambda_{j1}\xi_{i1} + \lambda_{j2}\xi_{i2} + \delta_{ij}$$

where the score of variable j for individual i is a function of intercept τ for variable j , factor loadings λ_{j1} and λ_{j2} of factors ξ_{i1} and ξ_{i2} respectively, and a residual term, δ_{ij} (Bollen, 2009; Kaplan, 2008; Kline, 2015).

Several modifications to the CFA models are made to estimate LGM. Since LGM estimates a latent mean structure in addition to the latent variance/covariance structure, the number of parameters estimated must not exceed $v(v + 3)/2$, v being the number of endogenous variables (Kline, 2015). To meet that rule, the values in the intercept vector are set to zeros for model identification (Preacher et al., 2008). Second, factor loadings are fixed to reflect the hypothesized pattern of growth over a certain period. Doing so results in factor parameters representing average and variability in the intercept and rate(s) of growth. The residual variance/covariance structure can be

manipulated to reflect autocorrelation due to time series data, such as autoregressive 1 (AR1; Preacher et al., 2008).

For illustrative purposes, let us consider a linear LGM with six measurement occasions (Figure 2). The framework of linear LGM with six measurement occasions can be represented by in matrix form as:

$$X = \Lambda \xi + \delta$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

where $X_1 - X_6$ represent the six measurement occasions, Λ represents the factor loadings matrix, which are constrained to reflect linear pattern of growth, and ξ is the latent variables vector, which represent the intercept and slope. The model implied variance/covariance matrix is the same as one illustrated above for CFA. In addition to the variance/covariance matrix, a mean structure is estimated for the latent variables (α_1, α_1). The model can also be represented in linear form as a function of factor means:

$$X_{it} = \xi_{i1} + \lambda_2 \xi_{i2} + \delta_{it}$$

$$\xi_{i1} = \alpha_1 + \zeta_{1i}$$

$$\xi_{i2} = \alpha_2 + \zeta_{2i}$$

where X_{it} is the observed score for person i at time t ; λ_2 is the factor loadings for the slope factor; and δ_{ij} is the residual for person i at time t . α_1 is the estimated average intercept across all individuals, at the initial measurement occasion. ζ_{1i} is the deviation

of person i 's estimated intercept from the average intercept. α_2 is the average estimated linear slope for the measure across all time points and all individuals; ζ_{2i} is the deviation of person's i 's estimated linear slope from the average slope (Figure 2; Preacher et al., 2008).

LGM has many extensions beyond the linear growth trajectory example illustrated. Within the LGM framework, complex growth trajectories such as polynomial, piecewise, and unspecified patterns can be specified. Growth patterns other than linear, would have a different factor structure compared to the example illustrated (Flora, 2008; Kline, 2015; Preacher et al., 2008). Additionally, within the LGM framework, complex designs such as cohort sequential, group comparisons, parallel processes, multilevel models, higher order models, and individually varying measurement occasions, can be specified.

LGM has several advantages over other statistical approaches to estimating growth trajectories, such as repeated measures analysis of variance (ANOVA), ordinary least squares (OLS) regression, and hierarchical linear modeling (HLM). Compared to repeated measures ANOVA and OLS, LGM allows for the modeling of complex residual/variance covariance structure, as well as allows growth terms to be predicted by other variables in the model (Duncan & Duncan, 2009; Muthén & Curran, 1997). Although HLM can achieve the same advantages listed above, HLM does not provide indices of model fit (other than of R^2), and does not allow the estimation of latent variables or growth factors to serve as predictors of other variables in the same model (Geiser, 2012; Kline, 2015; Preacher et al., 2008).

There are several assumptions that need to be met for LGM estimation, which can be divided into two categories: model specific and estimation specific. The first model specific assumption posits that LGM models must be estimated with a large sample size. SEM is a large sample size approach and LGM is no exception. Several recommendations exist for determining sample size for SEM based models, including the $N: q$ rule, which requires at least 20 participants (N) for each parameter (q) estimated, or no less than 200 cases (Kline, 2015). Power analysis is a more appropriate method of determining appropriate sample size for the model one wishes to estimate and is becoming more widely used (Hancock & Mueller, 2013; MacCallum, Browne, & Sugawara, 1996; Zhang & Wang, 2009).

The second assumption applies to the number of measurement occasions needed. The $M + 2$ rule needs to be followed when determining the minimum number of measurement occasions necessary to estimate a certain growth pattern, with M being the power of the growth function. For example, for a linear slope ($M = 1$) to be estimated, at least three measurement occasions are needed, and for a quadratic growth ($M = 2$), at least four measurement occasions are required (Cohen, Cohen, West, & Aiken, 2013; Preacher et al., 2008).

LGM assumes that all individuals at each measurement occasion are measured at the same time (Mehta & West, 2000). However, advancements in LGM have allowed for some relaxation of that assumption. More complex LGM models that account for differences in groups that have different measurement schedules can be estimated (Allison, 1987; Duncan, Duncan, 2004; Muthén, Kaplan, & Hollis, 1987). If there is a

limited number of measurement schedules, then a multiple group analyses can be estimated by placing individuals with the same measurement schedule in the same group (Preacher et al., 2008). Additionally, some SEM software, such as Mplus¹, can estimate individual slope loadings that account for the fact that not all individuals have the same measurement schedule (Muthén & Muthén, 2016).

It is also assumed that endogenous variables are measured without error. However, LGM allows for specifying second order factor models that permit the modeling of growth in latent variables while accounting for measurement error (Preacher et al., 2008). Lastly, the assumption of longitudinal invariance must be made for interpretation of results to be meaningful.

Maximum Likelihood Estimation

Most SEM models in educational research are estimated using normal based MLE (*nMLE*). *nMLE* has been built into commonly used statistical software, such as Mplus and Lisrel, is less computationally compared to other estimation methods², and is generally robust to minor violations of assumptions (Eliason, 1993). For a correctly specified model, when assumptions are met, *nMLE* estimates are said to be (Bollen, 2009):

¹ Using the “TSCORE” option under ‘variables’ command.

² Such as Bayesian estimation, which can be used to accommodate non-normal data but is very computationally taxing and time intensive (e.g. Tong & Zhang, 2014; Zhang et al., 2013) and weighted least squares estimator (WLS), which requires a large weight matrix (Hancock & Mueller, 2013).

1. *Consistent* – as sample size increases, the discrepancy between the parameter estimates derived from MLE and the population values decreases, or does not become greater than a small value
2. *Asymptotically unbiased* – in large samples, the parameter estimates derived from n MLE are equal to population values
3. *Asymptotically efficient* - compared to other estimators, MLE achieves low standard errors, and thus less variance and more confidence in parameter estimates

Conceptually, MLE is a class of algorithms that yield population parameters estimates under which the observed sample data are most likely to occur. MLE proceeds by maximizing a likelihood function, which involves an iterative process that begins plugging in reasonable starting values for each parameter and continues changing the values each iteration. At the end of the process, if the model converges and maxima is reached, the parameter estimates that maximized the likelihood function are reported. Non-convergence occurs when no maxima is reached after a certain pre-specified number of iterations (Eliason, 1993; Hox, Moerbeek, & Van de Schoot, 2010).

In the SEM framework, the fitting function that is maximized is³:

$$F_{ML}(\theta) = \log|\Sigma(\theta)| + \text{tr}(S\Sigma^{-1}(\theta)) - \log|S| - (p + q)$$

³ $S = E[(y - E(y))(y - E(y))']$; $E(y) = \mu$ replaced with $E(y) = \Lambda\mu\xi$ to estimate latent mean structure (Bentler & Yuan, 2000)

Where θ is the parameter to be estimates, $\Sigma(\theta)$ is the specified structural parameter matrix, S is the sample variance/covariance matrix, p is the number of exogenous variables, and q is the number of endogenous variables (Bollen, 2009). In essence, by maximizing the likelihood function, the final set of parameters minimizes the discrepancy between the observed sample covariance matrix and the model implied covariance matrix (Geiser, 2012).

The following is a list of the major assumptions for n MLE estimation of SEM based models (Bollen, 2009; Kline, 2015):

1. Large sample size
2. Variance/covariance matrix is positive definite
3. Observations for different subjects are independent
4. Model is correctly specified
5. Endogenous variables are continuous and follow a multivariate normal distribution
6. Latent variables are continuous and follow a multivariate normal distribution

Assumptions about the distribution of the endogenous variables and latent variables need to be made in order to derive the estimates that maximize the n MLE likelihood function. However, the general framework of MLE is very flexible and can accommodate other distributions. The normal distribution is simply imposed in common applications of MLE (Eliason, 1993).

Full Information Maximum Likelihood and Missing Data

Using the *n*MLE framework, the most common approach to handle missing data is to apply full information maximum likelihood (FIML). FIML is more popular in SEM compared to other methods (e.g. multiple imputation or listwise deletion) because it generally outperforms other methods (Enders, 2001; Li & Lomax, 2017) especially when fitting LGM (Shin, Davidson, & Long, 2009). FIML treats missing observations as random variables that are removed from the likelihood function, as if they were never sampled.

Assuming that data are missing completely at random (MCAR) or missing at random (MAR), FIML allows the calculation and maximization of the likelihood function by using all the available data from the variables being modeled, without imputing missing values. This approach when implemented within *n*MLE relies on the same distributional assumptions mentioned earlier; thus, it can only be guaranteed to provide estimates that are efficient and unbiased when distributional assumptions hold (Bollen, 2009; Kaplan, 2008).

Effects of Non-normality on *n*MLE Results

When any of the assumptions mentioned earlier are violated, the *n*MLE parameter estimates and standard errors can be biased and the model fit statistics less accurate. Generally, the greater the non-normality, the greater the impact on the results. The consequences of the violation of the normality assumption are augmented when there are missing data (Yuan & Bentler, 2001).

Model Fit. In regards to model fit indices, when the endogenous variables are leptokurtic, Chi-square values tend to be inflated, and when platykurtic⁴, Chi-square values tend to be deflated (Hu, Bentler, & Kano, 1992; Kline, 2015). Inflated chi-square values would lead to over-rejection of correctly specified models (e.g. Enders, 2001; Powell & Shafer, 2001; Ryu, 2011). Additionally, outliers have been shown to result in misleading and contradictory fit indices (e.g. Yuan & Bentler, 2001; Yuan & Zhong, 2013), especially when coupled with missingness (Lei & Lomax, 2005). Although *n*MLE is robust to minor violations of normality, Savalie (2008) found that if missing data percentage is greater than 10%, the robustness of *n*MLE Chi-Square to even mild violation of normality could not be guaranteed.

Standard Errors. The effects of non-normally distributed endogenous variables on the estimation of standard errors is well documented in the literature. Non-normality, especially leptokurtic, can sometime lead to underestimation of standard errors, and thus inflation of Type I error. Other times, especially when the data are platykurtic, it can lead to overestimation of standard errors and inflation of Type II error/reduction in power (e.g. Bollen, 2009; Enders, 2001; Hox, 2010; Kaplan, 2008; Kline, 2015, Yuan & Chan, 2005).

Parameter Estimates. Furthermore, when the normality assumption is not met, the parameter estimates in SEM are also affected. The impact of non-normality on parameter estimated has been less researched. There is a misconception that *n*MLE

⁴ Observed data is rarely platykurtic (Salvei, 2014)

parameter estimates are robust to violations of normality assumptions; however, there is sufficient evidence to suggest otherwise. Yuan, Bentler and Chan (2004) and Zu and Yuan (2010) found that longer-than-normal tails or outliers resulted in unreliable SEM parameter estimates. Similarly, Yuan and Zhong (2013) found that even a small percentage of outliers significantly biased parameter estimates. When nonnormality is accompanied by missing data, the consequences of the violation to the distributional assumptions are compounded. Shin, Davidson and Long (2009) found that when data are non-normal, missing data, even when MCAR, resulted in biased LGM parameter estimates, specifically for the variance of growth factors.

Bias in parameter estimates may be especially problematic for LGM, since parameter estimates include not only a variance/covariance structure, but also a mean structure that reflects average growth. The consequences can be severe when considering that LGM is widely used in the educational field to quantify typical growth on educational outcomes. If the true distributions of latent growth factors were non-normally distributed, then even if the mean is correctly estimated, its interpretation would be misleading, as the mean is no longer the best estimate of central tendency. However, if skewness parameters for the latent growth factors are also estimated, then the mean can be interpreted in the context of the shape of the distributions.

Robust Modifications

Several robust modifications that remedy the consequences of the violation of the normality assumption exist. The robust approaches can be divided into two main categories:

1. Normalization

2. Corrections

Normalization. Transformations and the removal of outliers are two ways of normalizing data. Transformations involve applying a mathematical function to the observed data in order to change the shape of the distribution to a better approximation of a normal distribution, while retaining information about the rank order of individuals and their variability (Cohen et al., 2013). Many examples of transformations for regression models, path models and SEM models exist in literature (e.g. Breiman & Friedman, 1985; Montfort, Mooijaart, & Meijerink, 2009). Transformations may represent a statistical solution but have serious practical limitations. The interpretation of the transformed variables is often very challenging, and results in very subjective conclusions. That is especially problematic for LGM, where the unit of measurement of the observed variables is meaningful in educational research.

Removal of outliers, or extreme values, which are often the cause for non-normality, can also normalize the observed data. However, great disagreement exists on what the definition of an outlier ought to be. Some regard outliers to be any observation that's more than three standard deviations away from the mean or a Mahalanobis distance greater than a predetermine value (e.g. Cohen et al., 2013). Any removal of data results in loss of information and reduction in power. In longitudinal research, where attrition is common, removing additional observations simply because they are considered outliers or extremes does not provide a good statistical or practical solution. Unless observations are removed due to data entry error or other errors, removing

outliers results in risking removal of information about the population from which the sample was obtained.

Corrections. There is a variety of different approaches to robust corrections within n MLE, and almost all are constructed using the same rationale: correct for inefficiency of the estimator (Savalei & Falk, 2014). Robust n MLE (n MLR) is the most common approach due to its availability in commercial software, and superiority to other corrections (e.g. Curran, West, & Finch, 1996; Gold, Bentler & Kim, 2003). Satorra and Bentler (1998) proposed n MLR as a method to correct chi-square statistics and standard errors by a degree proportionate to the extent of the multivariate kurtosis of the observed data.

The purpose of n MLR is to compute standard errors and chi-square-based model fit indices that don't rely on the assumption of multivariate normality. Standard errors are computed using a sandwich estimator, where a variation of naïve standard errors form the “bread” and an asymptotically distribution free correction (ADF; Browne, 1984) is the meat (Savalei, 2008). As for chi-square computation, n MLR adjusts the likelihood ratio test so that it's mean coincides with the mean of a chi-square distribution with appropriate degrees of freedom (Asparouhov & Muthen, 2005)⁵. There are several limitations of n MLR. First, it only considers multivariate kurtosis and not skewness.

⁵ Other algorithms of robust n MLE exist, such as MLM and MLMV (don't differ in SE but differ in Chi-square). MLM and MLMV based on Satorra & Bentler (1994) not available with missing data in Mplus

Second, parameter estimates are not adjusted or corrected. Third, n MLR's effectiveness depends on complete data (Savalei, 2008).

Other less common robust modifications include methods that use Huber Type weights to down weigh observations that are less likely to occur (Yuan and Bentler, 1998; Yuan & Zhang, 2012; Zhong & Yuan, 2011). The drawbacks of weighting approaches are the need to select the proportion of observations that need to be down weighted and the difficulty in obtaining likelihood-based information criteria for model selection (Lai & Zhang, 2017). Some proposed modifications consider missing data, such as 2-stage procedures (Tong, 2014; Yuan & Zhang, 2012), which first obtains saturated estimates of the population means and covariance matrix and then adjust standard errors and test based on amount of missing data, to reflect uncertainty (Savalei & Falk, 2014). These approaches are much less widely used compared to n MLR in the estimation of SEM, evidence that n MLR outperforms these methods (Zhong & Yuan, 2011), and a lack of easy-to-use software to carry them out.

Multivariate Skew t Distribution

The multivariate skew t maximum likelihood approach (st MLE) may potentially overcome many of the limitations mentioned in the previous section. st MLE can be used to not only provide robust estimates but increase the amount of information extracted from the observed data to estimate the model. Using st MLE also allows for the utilization of FIML without the normality distributional assumption. Using the correct distributional assumptions for the data, even with the presence of missing data, improves accuracy of the model (Asparouhov & Muthén, 2016).

Conceptually, when using *st*MLE to estimate LGM, the skewness and kurtosis of the multivariate distribution of the endogenous variables are incorporated into the likelihood function, and the latent variables are assumed to follow a multivariate skew *t* distribution as shown below:

$$Y \sim tMST(\mu, \Sigma, \delta, \nu)$$

where *Y* is the latent variable, μ is a vector of means, Σ is a variance/covariance matrix of the latent variables, and δ is a vector of skew parameters for each latent variable, and ν is a positive ‘degrees of freedom’ parameter’ (kurtosis) for the multivariate distribution of latent variables. The ν parameter can be interpreted as how much thicker the tails of the multivariate distribution are compared to a normal multivariate distribution⁶. Similarly, δ is interpreted as the amount of skewness, positive or negative, for the univariate distribution of each latent variable.

*st*MLE can accommodate, normal, skew normal, and *t* distributed endogenous and latent variables, as they are all considered special forms of the skew *t* distribution. Fixing the skewness parameter, δ , to zero gives the *t*-distribution. Fixing the degrees of freedom parameters ν at a large value produces a skew-normal distribution. Fixing skewness to zero and degrees of freedom to a large value produces the normal distribution (Muthén & Asparouhov, 2015).

⁶ Skew *t* distribution can be used to model any level of kurtosis (Asparouhov & Muthén, 2016).

The application of *st*MLE to the estimation of SEM is similar to *n*MLE in the sense that the distribution of the covariates is not modeled nor considered (Muthén & Asparouhov, 2015). When using *st*MLE to estimate LGM, the same identification rules discussed earlier apply. In addition to mean, variances, and covariances, distributional information is also extracted in order to estimate the skewness of each latent variable (including residuals) and the degrees of freedom of the multivariate distribution. If only the intercept and slope(s) factors skewness parameters are significant, then the skewness of the endogenous variables is fully explained by the growth parameters. If residuals are also skewed, it would indicate that some individuals asymmetrically deviate from their growth trajectory.

The application of *st*MLE to the estimation of SEM models has become more feasible with the incorporation of the “DISTRIBUTION” option in Mplus (available only with TYPE = GENERAL and MIXTURE; Muthén & Muthén, 2016). Using the “DISTRIBUTION = SKEWT” option, allows for the estimation of factor skewness and degree of freedom parameters. The standard error estimates are based on the inverse of the information matrix and computed using the sandwich estimation and direct maximization (Asparouhov & Muthén, 2016). In Mplus, model fit indices can be requested by specifying “H1MODEL” in “OUTPUT”⁷. The output provides skewness for each latent variable, and skewness for residuals can be requested by including {Y} under “MODEL”.

⁷ Does not provide SRMR in Mplus 8.

Examination and Application of *st*MLE

Multivariate skew t distribution has been extensively discussed in the literature (e.g. Lee et al., 2016; Lin et al., 2014), but it has not been until recently that it has been gaining more attention in SEM framework, due to the recent implementation of the method in Mplus. Muthén and Asparouhov (2015), and Asparouhov and Muthén (2016) demonstrated the benefits of using it with non-normal data in Growth Mixture Modeling (GMM) analyses. It was found that *st*MLE was more effective at extracting the correct number of classes in growth mixture modeling (GMM) with non-normal data compared to *n*MLE, given large sample size and adequate number of random starts (Asparouhov & Muthén, 2016; Muthén & Asparouhov, 2015). *st*MLE's performance was also examined with path models with covariates, and factor models, and it was found that parameter bias values were all less than 0.1 and coverage values close to 95% (Asparouhov & Muthén, 2016). Additionally, Lai and Zhang (2017) explored the performance of CFA fit indices using t based MLE in Mplus, in the presence of outliers, and found that they were relatively stable with sample sizes above 200. Specifically, Lai and Zhang (2017) found that outliers resulted in inaccurate factor model fit and lower convergence rates when estimated with *n*MLE; however, *t*MLE decreased the impact of outliers when sample size was large.

On the other hand, Hohmann, Holtmann, and Eid (2018) found acceptable performance of *st*MLE when sample size was large but parameter estimates were somewhat biased. There are other potential drawbacks to using *st*MLE. *st*MLE requires the extraction of more information from the observed data and more parameters are

estimated. Therefore, a larger sample size may be required for the analysis to have sufficient power, especially for distributional parameters (Asparouhov & Muthén, 2016). A second potential drawback is that the skew t model may yield biased parameter estimates when the endogenous variables and latent variables are in fact normal. Recent research in t -based Bayesian estimation of LGM indicated that when the latent variables are normally distributed, using t -based estimation results in biased parameter estimates, especially with small sample sizes (Tong & Zhang, 2012; Zhang, Lai & Tong, 2013).

As for model comparison, it may be problematic when the st MLE model is compared to n MLE, due to the degrees of freedom parameter, which approaches positive infinity for a normal distribution. The null hypothesis that the degrees of freedom parameter equals infinity is not possible to test, and $1/df = 0$ serves as an approximation, but not much research exists on whether it is a sufficient approximation⁸.

Gaps in the Literature and Purpose of the Study

Based on the application of st MLE option in Mplus to GMM, path models, and factor models, there is evidence to suggest that it may operate well in estimating LGM parameters as well. Yet, there is also evidence of potential drawbacks. No studies have investigated the potential benefits (and drawbacks) of using st MLE over n MLR in the estimation of LGM under different distributional and missing data conditions.

Theoretically, with sufficient sample size, st MLE would provide more reliable parameter estimates when the normality assumption is violated. This investigated the

⁸ Asparouhov and Muthén (2016) also discuss the $\lambda = 0$, which appears only in real and not simulated data.

performance of the *st*MLE option in Mplus in the estimation of LGM growth parameters. First, *st*MLE was used to estimate LGM for the Peabody Individual Achievement Test math subset, with data from the National Longitudinal Survey of Youth, 1997 Cohort, to demonstrate its application to real educational data. Second, the performance of *st*MLE was investigated through a Monte Carlo simulation.

2. METHOD: EMPIRICAL EXAMPLE

Participants and Data Structure

A subset of the National Longitudinal Survey of Youth 1997 Cohort (NLS97; Bureau of Labor Statistics, U.S. Department of Labor, 2005) was used to demonstrate the application of *st*MLE to real data and show the differences between *st*MLE. The NLSY97 is made up of data from 8984 individuals who were born between 1990 and 1984 and were 12-17 years old when first interviewed in 1997. The sample consisted of approximately 52% non-black/non-Hispanic, 26% black, 21% Hispanic or Latino/Latina, and less than 1% mixed race (Cooksey, 2018).

Specifically, longitudinal data on the Peabody Individual Achievement Test (PIAT; User's Guide-Assessments-PIAT Math) mathematics measure, which were available yearly, from 1997-2001, were used for the analyses. The chosen subset consisted PIAT mathematics scores during the years 1997-2001 for participants who were born in 1984 (12-13 years old in 1997). Only individuals who had scores for at least two time points were included in the analysis. The resulting subset included 1716 students.

Graphical and quantitative checks of normality indicated that all five time points were non-normally distributed (Table 1; Figure 3). Data appear to be negatively skewed for all time points, and slightly leptokurtic. Shapiro-Wilk and Kolmogorov-Smirnov tests of univariate normality confirmed that the data are significantly non-normally distributed (Table 2). Coverage ranged from approximately 76% to 95%, with each time point having approximately 15-25% missing observations (Table 3).

Model Specification and Analysis

For simplicity, the illustration ignored the multilevel structure of the data. Additionally, no sampling weights were applied⁹. Based on previous research (e.g. Tong & Zhang, 2012; Zhang et al., 2013), and visual examination of the individual growth plots (Figure 4), a linear growth pattern was specified for the PIAT. The growth in PIAT scores over five measurement occasions was modeled as:

$$PIAT_{it} = \xi_{1i} + \lambda_2 \xi_{2i} + \delta_{it}$$

$$\xi_{1i} = \alpha_1 + \zeta_{1i}$$

$$\xi_{2i} = \alpha_2 + \zeta_{2i}$$

Where ξ_{1i} is the intercept for person i , ξ_{2i} is the slope for person i , δ_{it} is the residual term for person i at time t . α_1 and α_2 represent the means and ζ_{1i} and ζ_{2i} represents the deviation from the means for person i , for the intercept and the slope factors respectively. Additionally, the correlation between the intercept and slope was estimated. Heteroscedastic and independent residual structure was specified. When using *st*MLE, the residual terms were assumed to be non-skewed (i.e. source of skewness is intercept and slope factors).

The data were analyzed in Mplus under two estimation procedures: *n*MLR and *st*MLE. Since both estimation procedures use Maximum Likelihood, the “ESTIMATOR” option in Mplus was set to “MLR”. However, in order to change the distribution on which the estimation was based, “SKEWT” was set for the

⁹ The Skew t distribution option in Mplus is not available with TYPE = COMPLEX (only GENERAL and MIXTURE).

“DISTRIBUTION” option in Mplus, where normal distribution is the default (Muthén & Muthén, 2016). Additionally, the “STARTS” option was used to specify the number of random sets of starting values to generate in the initial stage (100) and the number of optimizations to use in the final stage (20), as recommended by Muthén and Muthén (2016).

The “DSITRIBUTION” option, which has been implemented to Mplus version 7.2 and later (Muthén & Muthén, 2016) has several options: “NORMAL”, “SKEWNORMAL”, “TDISTRIBUTION”, and “SKEWT”. This option can only be used in conjunction with continuous observed variables and continuous factors and cannot be used with models that require numerical integration (Asparouhov & Muthén, 2016; Muthén & Asparouhov, 2015). When the “DISTRIBUTION” option is set to “SKEWT”, a skew parameter is estimated for each endogenous and latent variable, and one degrees of freedom parameter is estimated for the multivariate distribution of latent variables.

3. RESULTS: EMPIRICAL EXAMPLE

The results of the empirical example indicate that one might make different conclusions about students' growth on PIAT mathematics for the subsample of the NLSY97 depending on the analysis used (Table 4). When using *n*MLR, the linear model fit the data well, and the results indicated there was significant variability in the intercept ($\phi_{11} = 207.08$, $p < .05$) and the slope ($\phi_{22} = 2.65$, $p < .05$) factors. The mean intercept was $\alpha_1 = 68.65$, $p < .05$ and the mean slope was $\alpha_2 = 2.89$, $p < .05$. The *n*MLR results indicate that at the first time point, students are predicted to have score of approximately 69, and grow at a rate of almost 3 points per year on average. Furthermore, the covariance between the slope and intercept was not significant ($\phi_{12} = -1.90$, n.s). Lastly, approximately 98% of the variability in the endogenous variables was explained by the model.

Compared to *n*MLR, *st*MLE model fit could not be computed due to variance/covariance matrix not being positive definite. Fixing parameters did not resolve the error. It is possible that the issue was due to the $\lambda = 0$ problem, previously discussed¹⁰. *st*MLE resulted in smaller residual variances, indicating that the model explained approximately 99% of the variability in the PIAT math scores at each time point. The variability of the intercept ($\phi_{11} = 44.08$, $p < .05$) and the slope ($\phi_{22} = 1.41$, $p < .05$) factors were smaller, and the means were slightly higher ($\alpha_1 = 87.37$, $p < .05$; $\alpha_2 = 3.25$, $p < .05$). Unlike the *n*MLR results, *st*MLE results indicated that the slope and

¹⁰For this illustration, λ values were below .01.

intercept are significantly correlated ($\phi_{12} = -9.77, p < .05$; Figure 5). The *st*MLE approach also provided additional information that was not provided with the *n*MLR estimation: the skewness of the latent variables. Results indicated that the intercept was highly negatively skewed (skewness = $-17.01, p < .05$), and the slope factor was slightly but significantly negatively skewed (skewness = $-0.79, p < .05$; Figure 6). The degrees of freedom of the multivariate distribution of latent variables was $3.14, p < .05$.

It is possible to compare the *st*MLE and the *n*MLR models to determine which one fits the data best; however, those methods have not been widely used or thoroughly investigated (Asparouhov & Muthén, 2016). As recommended by Asparouhov and Muthén (2016), models were compared with the Bayesian information criterion (BIC; Schwarz, 1978)¹¹. For this scenario, the results of BIC comparison indicated that the *st*MLE model fit the data better than the *n*MLE model, with the *st*MLE model having a smaller BIC value.

¹¹ LRT results can be questionable, because the null is that the degrees of freedom parameter = infinity is not possible. $1/df = 0$ is an approximate null, but not much research exists on whether it is a sufficient approximation.

4. METHOD: MONTE CARLO SIMULATION

Population Growth Model

In order to determine the quality of *st*MLE parameter estimates, compared to *n*MLR, a Monte Carlo simulation study was conducted. Two population models were specified: ‘Linear Growth’ and ‘Zero Growth’ models. For both models, population values were based in part on parameter estimates from the empirical study discussed above. Data was generated based on a linear LGM, with five time points, and heteroscedastic and independent residuals. The “Linear Growth” population data was generated based on the following matrices:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta;$$

$$\alpha_1 = 60;$$

$$\alpha_2 = 3;$$

$$\Phi = \begin{bmatrix} 40 & -5 \\ -5 & 2 \end{bmatrix} \quad \Theta = \begin{bmatrix} 40 & 0 & 0 & 0 & 0 \\ 0 & 35 & 0 & 0 & 0 \\ 0 & 0 & 30 & 0 & 0 \\ 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0 & 15 \end{bmatrix}$$

As for the zero-growth model, the same matrices above were used; however, the slope factor mean and the correlation between intercept and slope were set to zero. Data was generated using the package SIMSEM in the statistical software R (Pornprasertmanit, Miller & Schoemann, 2016).

Simulation Conditions

Five simulation factors were manipulated: population model, sample size, distribution of latent variables, percentage of attrition/missingness, and estimation

procedure (Table 5). The two population models are described above, and were chosen so that both power and Type I error for each estimation procedure can be examined. Model specification was fixed across all conditions. Five hundred replications were analyzed per condition. Overall, there were 72 simulation conditions, or 36 conditions per estimation procedure (18,000 replications total).

Sample Size. Sample sizes were set to 100, 500, and 1000. The minimum was chosen as 100 since most SEM models are estimated with 100+ observations (Jackson et al., 2009) and SEM is considered a large samples approach (Kline, 2015). Additionally, the max sample size was 1000 because *st*MLE may require larger sample size to converge, especially when there are missing data, and 1000 would be considered large in several educational research fields.

Distribution. The degree of non-normality of the endogenous variables was manipulated indirectly by specifying the distribution of all latent variables (intercept factor, slope factor, and residuals). Specifically, a sequential method generates factor data first, based on specified distributions for the factors, then a set of equations is applied to obtain the observed data, following the Vale and Maurelli (1983) method (Pornprasertmanit et al., 2016). With that approach, if all latent variables are normally distributed, then the endogenous variables will in turn also be normally distributed, since they are the sum of a linear combination of latent variables.

The distribution of the latent variables was manipulated into three conditions: normal, moderate non-normality, and severe non-normality. Values for univariate skewness and Kurtosis for latent variables are based on Shin et al. (2009) and Cain,

Zhang, and Yuan (2017) review of typical univariate skewness and kurtosis values in real data. The following are the values used for different distributional conditions of all latent variables:

Normal: skewness = 0 and kurtosis = 3

Moderate non-normality: skewness = -3.25 and kurtosis = 5

Severe non-normality: skewness = -5 and kurtosis = 10

Missingness. In every longitudinal study, it is likely that some individuals will drop out of the study. All missingness in this study was assumed to be missing completely at random (MCAR). There were two missingness conditions specified: complete data and 10% attrition. For both conditions, the first time point does not have any missingness. Missingness was created by dropping 10% of the cases at each time point. Once a case has been dropped, it no longer had any data for the subsequent time points.

Estimation. Linear growth model was specified for all conditions, including the “Linear Growth” and the “No Growth” models. All models were correctly specified, and misspecification was not considered. All data was analyzed with two estimation approaches: *n*MLR and *st*MLE. All analyses are conducted in Mplus using the Monte Carlo option with population parameters specified as starting values¹². Extraction of simulation results were carried out in R, using the package “MplusAutomation” (Hallquist & Wiley, 2014).

¹² With the exception of distributional parameters (df and skewness) because specifying starting values resulted in non-convergence for the *st*MLE estimation procedure.

Evaluation Criteria

The performance of *n*MLR and *st*MLE was evaluated based convergence rates, relative bias, mean square error (MSE), 95% coverage, and Type I/Power. The accuracy of parameter estimates was evaluated through relative bias and MSE, and the performance of standard errors was evaluated with 95% coverage rates, Type/Power. First, convergence rate was determined by the proportion of replications that resulted in successful computation, with or without warnings¹³. Successful computations include those replications with warning messages. None of the successful computations with warning messages were excluded in order to maintain a sufficiently high number that is representative of replications across all conditions, an approach taken by other SEM researchers (e.g. Leite, 2007; Ulitzsch et al., 2017).

Second, relative bias refers to the difference between the averaged parameter estimates within each condition and the true population parameter (Ender's, 2001). Relative bias was calculated by the following formula (Muthén & Muthén, 2016):

$$Relative\ Bias = \frac{(\hat{\theta} - \theta)}{\theta} * 100$$

where $\hat{\theta}$ is average of parameter estimates, and θ is the population value. If the value of relative bias is below 15%, it is considered acceptable (Muthén, Kaplan, & Hollis, 1987).

¹³ Warning messages discussed separately

Third, MSE is interpreted as a combination of bias and sampling variability. It represents the overall accuracy of an estimate (Enders, 2001; Muthén & Muthén, 2016). When the parameter estimate is unbiased (absolute bias¹⁴ = 0), MSE represents efficiency, or the sampling variability, of a parameter estimate. MSE values for the same parameter from different conditions can be compared to determine the relative efficiency of estimation procedures. MSE is calculated as (Muthén & Muthén, 2016):

$$MSE(\hat{\theta}_j) = \frac{1}{N-1} \sum_{i=1}^N [\hat{\theta}_{ij} - \hat{\theta}_j]^2$$

where $\hat{\theta}_{ij}$ is the estimated parameter for sample i , $\hat{\theta}_j$ is the specified population value for parameter j , and N is the number of replications.

Fourth, coverage refers to the proportion of 95% confidence intervals (CI) that contained the true population value for each parameter (Muthén & Muthén, 2016). Coverage provides a measure of how well the parameters and their standard errors are estimated. It is expected that coverage values would be close to 0.95%. Lastly, power was determined for all parameters in the “linear Growth” conditions, and Type I error was determined for the Slope and Slope/Intercept covariance parameters for the “No Growth” conditions. Power is the probability of rejecting the null hypothesis when it is false, and is expected to be approximately 0.8. Type I error is, the probability of rejecting the null hypothesis when it is true, is expected to be below 0.05 (Muthén & Muthén, 2016).

¹⁴ Absolute bias is difference between average parameter estimate and population value

5. RESULTS: MONTE CARLO SIMULATION

Convergence Percentages

Convergence percentages were all 100% for n MLR across all conditions.

Convergence percentages were also approximately 100% for st MLE for conditions with $N \geq 500$. For conditions with $N = 100$, st MLE convergence percentages ranged from 91.6% to 98.8% for conditions with growth, and 85.8% to 98.2% for conditions with no growth. Specifically, convergence was lower when data was incomplete. Furthermore, st MLE convergence for conditions with $N = 100$ was lower for conditions with non-normally distributed growth factors, especially for conditions with zero growth (Table 6).

Although st MLE convergence percentages were greater than 85% for all conditions, a large proportion of those successful computations of conditions with non-normally distributed growth factors produced a warning message, and/or automatically fixed a random effects parameter to zero. The first warning message produced was:

“Warning: The latent variable covariance matrix (psi) in class 1 is not positive definite. This could indicate a negative variance/residual variance for a latent variable, a correlation greater or equal to one between two latent variables, or a linear dependency among more than two latent variables.”

Approximately 49.8% and 28.6% of replications, from growth and zero growth conditions respectively, produced warning message when growth factors were severely non-normal. Similarly, 64.2% and 81.18% of replications from growth and zero growth conditions respectively resulted in the mentioned warning message when growth factors

were moderately non-normal¹⁵. No warning messages were produced for conditions with normally distributed growth factors.

In regards to automatically fixed parameters, the following warning message was produced, sometimes in conjunction with the above message:

“One or more parameters were fixed to avoid singularity of the information matrix. The singularity is most likely because the model is not identified, or because of empty cells in the joint distribution of the categorical variables in the model.”

The parameters that were fixed as a result were the intercept factor variance, growth factors covariance, and in some cases one or more residual variance(s). Approximately 22.2% and 47.6% of replications, from growth and zero growth conditions respectively, resulted in some random effects being fixed to zero when growth factors were severely non-normal. Likewise, 7.8% and 3.2% of replications, from growth and zero growth conditions respectively, resulted in some random effects being fixed to zero when growth factors were moderately non-normal¹.

All replications that successfully converged were used to calculate bias, MSE, coverage, power and Type I error rates. Eliminating replications with warning messages would eliminate up to half of the replications for certain conditions, thus the entire distribution of parameters would not be considered (Leite, 2007).

¹⁵ Percentages based on conditions with $N = 1000$ and complete data. Percentages were higher when sample size was small, but did not differ across conditions with complete vs missing data.

Summary. Overall, *n*MLR convergence percentages were 100% cross all conditions. *st*MLE convergence percentages were at approximately 100% when sample size was high. When sample size was low, *st*MLE convergence percentages were lower when growth factors were non-normally distributed, there was missing data, and sample size was small. Although convergence percentages were generally high for *st*MLE, the majority of replications for conditions with growth factors that were non-normally distributed resulted in estimation issues such as warning messages and/or the automatic fixing of some of the random effects parameters. Estimation issues occurred more frequently for conditions with severely-non-normal data.

Parameter Relative Bias

In all almost all conditions, parameter bias was lower using *n*MLR estimation rather than *st*MLE. Across all conditions and parameters and with both estimation methods, conditions with 10% attrition did not notably differ in bias from conditions with complete data. Patterns in parameter bias for the growth factors variances, residual variances, and intercept factor mean were essentially identical across the conditions with growth and the conditions with no growth. Results for these parameters are presented for only the conditions with growth. All relative bias values presented are in percentages.

Random Effects. *n*MLR growth factor variance estimates were generally unbiased across all conditions, with bias values ranging from -0.767 to 1.017 for the intercept factor variance and from -1.185 to 1.330 for the slope factor variance. As for the growth factors covariance parameter, *n*MLR was unbiased only when data was normally distributed, with bias values ranging from -1.352 to 0.532. Growth factor

covariance in conditions with non-normally distributed growth factors was negatively biased, with average bias values of -71.534 for moderate non-normality and -38.033 for severe non-normality. Bias for random effects using *n*MLR did not systematically differ across conditions with missing or complete data, but decreased very slightly as sample size increased (Figure 5). When the growth factors covariance population value was set to zero (zero growth conditions), *n*MLR estimates were generally unbiased across all conditions (Figure 6). Lastly, *n*MLR estimates of residual variances were unbiased, average bias = -0.199 (Figure 5).

*st*MLE severely underestimated variances of growth factors with bias values ranging from approximately -24.401 to -96.187 for intercept factor variance, and -24.390 and -52.052 for slope factor variance. Intercept factor variance was more biased for conditions with non-normally distributed growth factors, with mean bias of -31.881 for normally distributed data, -89.781 for moderately non-normal data, and -91.161 for severely non-normal data. *st*MLE growth factors covariance parameter bias followed a similar pattern: parameter was less biased for condition with normally distributed growth factors, with bias values ranging from -52.158 to -25.940. For conditions with non-normally distributed growth factors, estimates were more severely negatively biased, with bias values ranging from -64.468 to -56.490 when data was moderately non-normal, and -73.310 to -70.144 when data was severely non-normal (Figure 5).

Since *st*MLE estimates of growth factor variances were underestimated, bias in correlations in addition to covariances between growth factors was also examined. Once

covariances have been standardized, results indicate that correlations were actually overestimated, with average bias values of 41.090 (Figure 5).

When the growth factors covariance population value was set to zero (zero growth conditions, *st*MLE estimates were unbiased only for conditions with normally distributed growth factors, with average estimates of 0.028 (Figure 6). For conditions with non-normally distributed growth factors, the slope factor mean parameter was underestimated, with estimated values ranging from -2.295 to -1.593 for conditions with moderately non-normal data, and -2.073 to -1.535 for conditions with severely non-normal data. Lastly, *st*MLE estimates were negatively biased across all conditions, ranging from -21.056 to -16.678, with no systematic differences were observed across conditions with different sample sizes, missingness, or growth factors distributions (Figure 5).

Fixed Effects. As for the fixed effects, *n*MLR produced unbiased estimates for the intercept factor mean across all conditions, ranging from -0.021 to 0.065. Similarly, the slope factor mean was generally unbiased for conditions with normal or severely non-normal growth factors, ranging from -0.920 to 0.080 (Figure 7). When growth factors were moderately non-normal, *n*MLR estimates of the slope factor mean was highly negatively biased (average bias = -71.534). No differences were observed across conditions with different sample sizes, or missingness (Figure 7). When the slope factor mean population value was set to zero, *n*MLR estimates were close to 0 for all conditions, ranging from -0.025 to 0.003 (Figure 8).

The bias of *st*MLE estimates of fixed effects followed a similar pattern; values were below 15% for all conditions except for the slope factor mean for conditions with moderately non-normal growth factors (Figure 7). Intercept factor mean was more accurately estimated for conditions with normally distributed growth factors (average bias = -0.146) compared to moderately non-normal (average bias = 11.210) or severely non-normal (average bias = 11.150). As for accuracy of slope factor estimates, for conditions with for conditions with normal or severely non-normal growth factors, slope factor mean was essentially unbiased, bias values ranging from 0.117 to 2.040 when data was normally distributed, and -5.097 to 6.033 when data was severely non-normal. However, when data was moderately non-normal, biased values ranged from -69.580 to -56.460; which is similar to what has been found with *n*MLR (Figure 7). When the slope factor mean population value was set to zero, *st*MLE parameter estimates were also fairly close to zero, with estimates ranging from -0.025 to 0.003 (Figure 8).

Distributional Parameters. The *st*MLE estimation method also estimated skewness parameters for the slope factor and intercept factors. Results indicated no differences across growth and zero growth conditions, or conditions with complete or missing data. Results for growth with complete data are illustrated and discussed (Figure 9). When growth factors were normally distributed, skewness for both intercept factor and Slope factor were very close to zero across all sample sizes (Figure 9). However, for conditions with non-normally distributed growth factors, intercept factor skewness was underestimated by -4.667 when data was moderately non-normal, and -2.926 when data was severely non-normal, on average. On the other hand, slope factor skewness was

overestimated by 2.900 when data was moderately non-normal, and 5.102 when data was severely non-normal, on average.

The degrees of freedom parameter, ν , was not estimated for most of the conditions with $N = 100$ (output indicated “NA” for average parameter estimates). Average ν values across $N = 500$ and $N = 1000$ conditions were 10.160 for conditions with normally distributed data, 9.709 for conditions with moderately non-normal data, and 9.708 for conditions with severely non-normal data. There were no differences in estimates across conditions with complete or missing data.

Summary. Overall, there evidence to suggest that non-normality of growth factor leads to negatively biased n MLR parameter estimates of the growth factors covariance when population value was not zero. Mixed evidence exists for the impact of non-normality of growth factors on n MLR slope factor mean estimate. st MLE estimates, on the other hand, resulted in negatively biased estimates for random effect parameters across almost all conditions. st MLE fixed effects estimates were generally accurate, with some variation depending on the conditions. As for distributional parameters, they were accurate only when growth factors were normally distributed. For conditions with non-normally distributed growth factors, skewness was greatly under or over-estimated. The population value for the degrees of freedom parameter was not specified¹⁶; however, the conditions with normal data should have a high value for degrees of freedom (Muthén &

¹⁶ Only univariate kurtosis was specified for data generation. Mplus does not provide an equation for the relationship between multivariate kurtosis and ν (Asparouhov & Muthén, 2016). The kurtosis for the univariate t distribution is $6/(\nu - 4)$

Asparouhov, 2015). The results indicate that degrees of freedom estimates did not differ across conditions with different distribution specifications. Overall, *n*MLR outperformed *st*MLE in almost all conditions. Surprisingly, *st*MLE seemed to worse for conditions with non-normal growth factors.

Mean Square Error of Parameter Estimates

Parameter accuracy was also evaluated with MSE, which is a metric of absolute bias (difference between average parameter estimate and population value) and efficiency. Similar to what was observed with bias, across all conditions and parameters and with both estimation methods, conditions with 10% attrition did not notably differ in MSE from conditions with complete data. Patterns in parameter MSE for the growth factors variances, residual variances, and intercept factor mean were essentially identical across the conditions with growth and the conditions with no growth. Results for these parameters are presented for only the conditions with growth.

Random Effects. Since growth factor variances' bias was essentially zero when using *n*MLR, MSE values represent efficiency. For the conditions with growth, *n*MLR growth factors variance estimates were most efficient for conditions with $N = 1000$. Specifically, *n*MLR intercept factor variance MSE values ranged from 86.045 to 105.239 for $N = 100$, 16.071 to 19.654 for $N = 500$, and 7.609 to 9.758 for $N = 1000$. *n*MLR slope factor variance MSE values followed the same pattern, but were much smaller in magnitude, with ranged from 0.604 to 1.059 for $N = 100$, 0.124 to 0.207 for $N = 500$, and 0.055 to 0.091 for $N = 1000$ (Figure 10).

As for the growth factor covariance parameter, after taking bias into account, the results indicate that *n*MLE estimates are efficient with large sample sizes regardless of distribution of growth factors, with MSE values ranging from 4.600 to 8.081 for conditions with $N = 100$, 0.939 to 13.932 for conditions with $N = 500$, and 0.444 to 13.389 and from to when $N = 1000$. When growth factors covariance population value was set to zero, parameter estimates were fairly consistent across all conditions, with mean $MSE = 2.522$ (Figure 11).

Lastly, residual variances were efficient when sample size was large, with MSE values averaging 36.010 for conditions with $N = 100$, 7.150 for conditions with $N = 500$, and 3.620 and from to when $N = 1000$ (Figure 10). For *n*MLR random effects parameters, conditions with complete data were slightly more efficient compared to conditions with 10% attrition; however, MSE did not differ across conditions with different growth factors distributions.

Since growth factor variances were negatively biased when using *st*MLE, MSE values represent a function of absolute bias and efficiency. Even after factoring out the effect of absolute bias, MSE values indicate that *st*MLE intercept factor variance is highly inefficient for conditions with non-normally distribute growth factors, MSE values ranging from 1196.783 to 1483.867. As for the slope factor variance, although biased, it was consistent across all conditions and replications (average $MSE = 0.885$; Table 10).

As for the growth factor covariance parameter, after taking bias into account, the results indicate that *st*MLE estimates are consistent with large sample sizes regardless of

distribution of growth factors, with MSE values ranging from 12.77 to 17.372 for conditions with $N = 100$, 3.518 to 14.481 for conditions with $N = 500$, and from 2.354 to 14.385 when $N = 1000$. When the growth factors covariance population value was set to zero, *stMLE* parameter estimates were efficient, with MSE values averaging at 4.693 (Figure 11). Lastly, residual variances parameters were efficient when sample size was large, with MSE average 63.833 for conditions with $N = 100$, 44.181 for conditions with $N = 500$, and 43.327 when $N = 1000$ (Figure 10). Efficiency did not systematically differ across conditions with different growth factors distributions. Conditions that have 10% attrition resulted in less efficient parameter estimates, especially when sample size was small.

Fixed Effects. *nMLR* intercept factor mean was efficient across all conditions, with MSE values ranging from 0.057 to 0.656. As for the slope factor mean estimates, MSE values were close to zero, ranging from 0.005 to 0.061. Conditions with higher sample size having slightly higher MSE (Figure 12). As for the slope factor mean estimates, MSE values for both estimation methods across all conditions were close to zero, ranging from 0.005 to 0.061 (Figure 13).

The efficiency *stMLE* estimates of fixed effects were somewhat mixed. After factoring out absolute bias, intercept factor estimates varied somewhat across replications, with MSE values ranging from 2.781 to 20.480 for conditions with normally distributed growth factors, 44.987 to 49.698 for conditions with moderately non-normal growth factors, and 44.044 to 51.214 for conditions with severely non-normal growth factors. Slope factor mean was efficient across all conditions, with MSE

ranging from 0.070 to 1.509 (Figure 12). Similarly, For the zero growth conditions, for which the slope factor population value was zero, MSE values were close to zero, ranging from 0.205 to 1.607 (Figure 13). Conditions with larger sample size resulted in higher parameter efficiency.

Summary. Overall, across both estimation methods, efficiency was higher when data was complete and sample size was high. *n*MLR parameter efficiency was generally not impacted by non-normality. *st*MLE estimates of random effects were consistent, except for intercept factor variance which was highly inefficient. This is most likely due to the automatic fixing of those parameters for a large percentage of the replications. As for the efficiency of *st*MLE of fixed effects, the slope factor mean was efficient, but the intercept factor mean varied somewhat across replications when data was non-normally distributed.

Coverage: 95% Confidence Interval

The percentage of 95% confidence intervals that contained the population values was examined. Conditions with 10% attrition did not differ in parameter coverage from conditions with complete data. Therefore, coverage results only for conditions with complete data are presented. Similarly, conditions with growth did not differ in parameter coverage from zero-growth for the growth factors variances, residual variances, and intercept factor mean. Results for these parameters are presented for only the conditions with growth. Coverage percentages for the covariance parameter and slope factor mean parameter when population value is set to zero is the same as 1-Type I error rate, therefore they are discussed in the next section.

Random Effects. *n*MLR's parameter coverage for slope factor and intercept factor variance was at approximately 95% or higher for all conditions, ranging from 94.6% to 99.6%, increasing with sample size (Table 7; Table 8). The coverage of the growth factors covariance parameter estimated by *n*MLR varied across conditions with different growth factors distributions. For conditions with normally distributed growth factors, coverage percentages ranged from 94.0% to 96.2%, with conditions with larger sample sizes having higher coverage. The trend is reversed, however, for conditions with non-normal growth factors. For conditions with moderately non-normal growth factors, coverage percentages ranged from 53.8% (for $N = 100$) to 0.0% ($N = 1000$). For conditions with severely non-normal growth factors, coverage percentages ranged from 78.0% (for $N = 100$) to 13.8% ($N = 1000$; Table 8). Lastly, for the residual variances, *n*MLR coverage percentages were above 95% for all conditions, ranging from 96.32% to 99.20%, with percentages being higher for conditions with larger sample size (Table 9).

When estimated with *st*MLE, intercept factor variance coverage was very low. Coverage percentages were worse for conditions with non-normal data. When growth factors were normally distributed, intercept factor coverage ranged from 16.4% to 56.1%. When growth factors were moderately non-normal, intercept factor variance coverage percentages ranged from 0.0% to 12.0%. When growth factors were severely non-normal, coverage percentages ranged from 0.0% to 12.9% (Table 7). Slope factor variance coverage was higher but did not reach 95% for any of the conditions (Table 8). Slope factor variance coverage followed the same pattern with coverage percentages ranging from 59.8% to 73.3% with normal data, 25.6% to 68.9% with moderately non-

normal data, and 32.0% to 63.0% with severely non-normal data (Table 8). Similarly, the growth covariance parameter's coverage average percentages were 72.7% with normal data, 22.8% with moderately non-normal data, and 19.37% with severely non-normal data (Table 9). Lastly *st*MLE residuals coverage was higher for conditions with smaller sample size, with coverage percentages ranging from 79.24 to 83.3 for $N = 100$, 35.75 to 40.88 for $N = 500$, and 13.88 to 16.04 for $N = 1000$ (Table 10). Coverage percentages were generally higher for conditions with normally distributed growth factors.

Fixed Effects. When estimated with *n*MLR, intercept factor mean coverage ranged from 94.60% to 96.0% (Table 10) and slope factor mean coverage ranged from 92.60 to 95.80 (Table 11), with percentages being higher for conditions with larger sample size. When estimated with *st*MLE, coverage percentages for fixed effects were below 95% for all conditions. Similar to what was found with random effects, coverage for fixed effects was better when data was normally distributed. For the intercept factor mean, when data was normally distributed, coverage increased as sample size increased, with percentages ranging from 56.1% to 90.8%. When data was moderately non-normal, the effect of sample size flipped; increase in sample size was associated with lower coverage. When data was moderately non-normal, coverage ranged from 11.7% to 0.0% (Table 11). For the slope factor mean, the effect of sample size was opposite to what was found with intercept factor mean. For conditions with normally distributed data, coverages ranged from 78.0% to 52.2%, with higher coverage being associated with lower sample size. When data was non-normal, coverage ranged from 64.4% to 89.6%

for moderately non-normal conditions, and 75.7% to 8.4% for severely non-normal conditions, with higher coverage associated with higher sample size (Table 12).

Summary. Overall, similar to previous results, non-normality seems to affect *n*MLR's performance only for growth factors covariance parameter when population value wasn't zero. *st*MLE coverage did not reach 95% for any of the conditions, but was generally higher for conditions with normally distributed growth factors. Furthermore, for most conditions, *st*MLE coverage percentages decreased as sample size increased.

Power

The power rates for each parameter across all conditions was assessed by examining the proportion of replications with significant parameter estimates, at the $p > .05$ level. Conditions with growth did not differ in power rates from zero-growth for the growth factors variances, residual variances, and intercept factor mean. Results for these parameters are presented for only the conditions with growth.

Random Effects. *n*MLR power rates for the intercept factor slope factor variance were high for conditions with $N = 500$ and $N = 1000$, ranging from .996 to 1.00. For conditions with $N = 100$, intercept factor variance power was still fairly high, with average power = 0.983, and slope factor variance power lower at average of 0.629. *n*MLR's estimates of the growth factors covariance parameter were inconsistently affected by non-normality. For conditions with normally distributed growth factors and severely-non-normal growth factors, power for the covariance parameter was high for $N = 500$ and $N = 1000$. However, for conditions with moderately non-normal data, power ranged from 0.098 to 0.560. As for average power for residual power was higher than

0.900 for all conditions (Figure 14). When sample size was small, power rates were slightly lower for conditions with non-normal growth factors and 10% attrition.

*st*MLE power rates for random effects parameter estimates were above 0.8 for most conditions with $N = 1000$ and complete data, with few exceptions. Power for intercept factor variance was high only when data was normally distributed. For conditions with moderately non-normal data, intercept factor variance power ranged from 0.252 to 0.580. For conditions with severely non-normal data, intercept factor variance power ranged from 0.278 to 0.559 (Figure 14).

Fixed Effects. Power of fixed effects parameter was above 0.8 for almost all conditions across both estimation methods. Power rates fell below 0.8 only for *st*MLE slope factor mean estimate for the condition with $N = 100$ and moderately non-normal growth factors (Figure 15).

Summary. When sample size was high, *n*MLR was sufficiently powered for all estimates, with the exception of the covariance parameter when data was moderately non-normal. Similarly, *st*MLE estimates were sufficiently powered for all parameters when sample size was high, except for the intercept factor variance and growth factors covariance, which varied depending on the condition. When data was non-normally distributed, *st*MLE's intercept factor variance estimates did not reach acceptable power. As for the growth factors covariance, power did not systematically differ across conditions. *st*MLE fixed effects estimates were sufficiently powered with large sample size.

Type I Error

Type I error did not differ across conditions with complete versus 10% attrition; therefore, only results for conditions with complete data are presented. For the no-growth condition, when estimated with *n*MLR, type I error rates for the growth factors covariance parameter was slightly above the nominal .05 level for conditions with non-normally distributed data. When data was normally distributed, Type I error rates were below .05 when sample size was high (Table 13). As for the slope factor mean parameter estimates, *n*MLR type I error rates were above .05 for almost all conditions, with highest rates being for conditions with small sample size and non-normally distributed growth factors (Table 14).

When estimated with *st*MLE, Type I error of both parameters was extremely high for most conditions. Type I error of the growth factors covariance parameter was acceptable only for conditional with $N \geq 500$ and normally distributed growth factors. However, for the other conditions, type I error rates average was 0.727 for conditions with moderately non-normal data, and 0.491 for conditions with severely non-normal data (Table 12). Type I error of the slope factor mean parameter, however, was above 0.05 for all conditions, with average rate being 0.208 for conditions with normal data, 0.727 for conditions with moderately non-normal data, and 0.491 for conditions with severely non-normal data (Table 14).

Overall, increase in sample size was associated with decrease in Type I error rates. Non-normality seems to slightly inflate Type I error rates of *n*MLR estimates.

*st*MLE estimates Type I error rates were above acceptable levels for almost all conditions, but were generally lower when growth factors were normally distributed.

6. DISCUSSION

Summary of Results

The development of methods that can accurately estimate population values based on non-normal data has been receiving more attention over the past decade, partially due to the advancements in statistical software. Skew t based maximum likelihood has been suggested several times in the SEM framework (e.g. Lee et al., 2016; Lin et al., 2014), but until recently, it was difficult to implement. With the introduction of the ‘DISTRIBUTION’ option in Mplus (Muthén & Muthén, 2016), applying st MLE to SEM has become more feasible. The purpose of this study was to examine the statistical performance of st MLE in estimating LGM under different growth factors distributional conditions. The performance was evaluated through application of the estimation method to analyzing data from the National Longitudinal Survey of Youth, as well as through a Monte Carlo simulation, comparing st MLE results with n MLR. The results of the empirical evaluation indicated that one would arrive at vastly different conclusions regarding the effect size and significance of the intercept factor variance, the growth factors covariance, and the residual variances, depending on whether the data is analyzed with n MLR or st MLE. The results of the simulation study indicated that, contrary to what was expected, st MLE estimates are less accurate compared to n MLR.

In regards to the effect of non-normality on normal based robust approaches, n MLR performed well despite violations to normality assumption even with the presence of missing data. Convergence rates were 100% for all conditions, Parameter estimates and standard errors functioned well for all parameters, except for growth factors

covariance and slope factor mean. When population value was not zero, the effect of non-normality on the growth factors covariance was fairly consistent. When data were non-normal, the covariance parameter was negatively biased, had low coverage rates, and low power rates. Slope factor mean was biased when data was moderately non-normal, but was otherwise unbiased and efficient across all other conditions.

As for *stMLE*'s performance, it was generally worse than *nMLR*. Although convergence rates were high, a large proportion of replications resulted in warning messages when data was non-normal. Warning messages indicated estimation issues such as the automatic fixing of several random effects parameters to zero. The most commonly fixed parameters were the intercept factor variance and the growth factors covariance, which greatly increased bias and inefficiency for those parameters. As for the remaining random effects, they were mostly negatively biased, but consistent. Fixed effects, on the other hand, were unbiased, but their standard errors did not perform well, as indicated by low coverage, power, and Type I error rates. *stMLE* did provide additional information about the skewness of the growth factors, but was accurate only when data was normally distributed. Across all conditions and all evaluation criteria, results indicate that *stMLE*'s performance is best when sample size was high, data was complete and was normally distributed.

Significance and Implications

There is mixed evidence in the literature about the impact of non-normality on parameter estimates in structural equation models. It has long been believed that when data is non-normally distributed, standard errors are inaccurate, but parameter estimates

remain unbiased. However, recent evidence suggests otherwise (e.g. Shin, et al., 2009; Zu & Yuan, 2010). The results of this study show that, while most parameters are unbiased regardless of the severity of the non-normality of growth factors, the growth covariance parameter is greatly affected by violations to normality assumption. This finding suggests that *n*MLR may not be appropriate to use with non-normal data when modeling more complex factor covariances in LGM, such as effects of time invariant covariates or parallel process models. Additionally, in this study, only the growth factors distributions were manipulated, while all residual terms were normally distributed. It may be the case that if residuals are non-normal, using *n*MLR would result in inaccurate estimates when modeling autocorrelation or adding time invariant covariates.

Unfortunately, *st*MLE option in Mplus to estimate LGM also has serious limitations under the conditions tested in this study. The results indicate warning messages should not be ignored, as they are most likely an indication of inaccurate solutions. Similar results were found in the application of *st*MLE to the estimation of mixture latent trait analysis, with a large proportion of random effects automatically fixed to zero, especially when sample size was small (Hohmann et al., 2018).

It is unknown whether excluding those replications that produced warning messages would result in better results. Hohmann et al. (2018) found acceptable performance for those replications without warning or error messages, when applying *st*MLE to mixture latent state-trait analysis, but parameter estimates were still somewhat biased across all conditions. It is unclear why the warning messages regarding the latent variable covariance matrix not being positive definite are produced; however, they seem

to occur more frequently when the non-normality was more severe, and did not occur at all what data was normally distributed. It is worth noting that these estimation issues occurred even after specifying the population value as the starting value. It is reasonable to expect even more estimation issues in empirical research, when starting values are not known.

Hohmann et al. (2018) hypothesized that these estimation issues may be occurring due to small population values of latent variances. However, in this study, population values for random effects were not close to zero. It may be the case that the maximum sample size in this study was still too low, given that distributional information needed to be extracted. When data was normally distributed, the degrees of freedom parameter was not estimated, thus reducing the final number of parameters estimated – which may be the reason why conditions with normally distributed data did not result in warning messages.

Additionally, the number of random starts used in this study may have been insufficient. The number of random starts chosen for this study was based on recommendation from the Mplus manual (Muthén & Muthén, 2016). However, in their studies Muthén and Asparouhov (2015) and Asparouhov and Muthén (2016) used a higher number of random starts when applying *st*MLE to growth mixture modeling. If *st*MLE requires large sample size ($N > 1000$) and a high number of random starts, two major disadvantages present themselves. First, obtaining such a large sample size in educational longitudinal research is often unfeasible. Second, *st*MLE is computationally

demanding and requires more time to converge. If the number of random starts was increased along with sample size, the computation time would be significantly higher.

Aside from parameter estimation issues, the performance of standard errors was not ideal as indicated by low coverage rates even for unbiased parameters, low power, and high type I error; which seems to be worse when data was severely non-normal. One potential reason is that standard errors are still based on the assumption that the sampling distribution is normal. Falk (2017) found that the sampling distributions take longer to approach normality when multivariate normality of endogenous variables is violated, especially for variance components. If that is the case, Bayesian methods have an advantage and should be further investigated as an alternative framework to implement skew t -based estimation (e.g. Tong & Zhang, 2014; Zhang et al., 2013).

Limitations and Future Research

The results should be interpreted in context of the following limitations. First, evaluation criteria were not calculated for replications with and without warning messages separately. The next step would be to evaluate whether replications without warning messages produce more accurate and reliable results. However, it is worth noting that even conditions without any warning messages still had issues with $stMLE$'s performance. Second, the effect of missingness was limited to 10% attrition, which was not sufficient to affect the performance of either estimation methods. Different missingness conditions should be tested (e.g. Shin et al, 2009). Third, growth factors distributions were manipulated by changing skewness and kurtosis values simultaneously; thus, the effects of skewness or kurtosis alone was not investigated.

Future research should investigate whether increase in sample size along with increase in number of random starts would improve estimation. Additionally, varying the source of non-normality (e.g. Ceiling/floor effects, outliers, and data contamination) would further provide information about the conditions under which *st*MLE presents an appropriate approach compared to other robust estimation approaches.

7. CONCLUSION

Using the *st*MLE option in Mplus as an alternative to *n*MLR may be beneficial under certain conditions. Based on the results of this study and previous studies, the option remains experimental, and output should be interpreted with caution. It appears that *st*MLE performs best with high sample size with data that is not severely non-normal and does not contain a large proportion of missing data. However, until further research sheds light on which conditions lead to trustworthy results for *st*MLE, other well-established robust methods, such as *n*MLR, may be more appropriate to use in substantive research.

REFERENCES

- Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological methodology*, 71-103.
- Asparouhov, T., & Muthén, B. (2005, November). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) research conference* (pp. 14-16).
- Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous non-normal skewed distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1), 1-19.
doi:10.1080/10705511.2014.947375
- Bollen, K. A. (2009). *Structural equations with latent variables*. Wiley.
- Brandt, H., & Klein, A. G. (2015). A Heterogeneous growth curve model for non-normal Data. *Multivariate behavioral research*, 50(4), 416-435.
- Breiman, L., & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391), 580-598.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37(1), 62-83.
- Bureau of Labor Statistics, US Department of Labor (2005). *Washington, DC*.

- Cain, M. K., Zhang, Z., & Yuan, K. H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior research methods*, 49(5), 1716-1735.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- Cooksey, E. C. (2018). Using the National Longitudinal Surveys of Youth (NLSY) to conduct life course analyses. In *Handbook of Life Course Health Development* (pp. 561-577). Springer, Cham.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological methods*, 1(1), 16.
- Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior therapy*, 35(2), 333-363.
- Duncan, T. E., & Duncan, S. C. (2009). The ABC's of LGM: an introductory guide to latent variable growth curve modeling. *Social and personality psychology compass*, 3(6), 979-991.
- Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice* (Vol. 96). Sage Publications.
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological methods*, 6(4), 352.

- Falk, C. F. (2017). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, 25(2), 244-266.
- Farley, D., Anderson, D., Irvin, P. S., & Tindal, G. (2017). Modeling reading growth in grades 3 to 5 with an alternate assessment. *Remedial and Special Education*, 38(4), 195-206.
- Flora, D. B. (2008). Specifying piecewise latent trajectory models for longitudinal data. *Structural Equation Modeling*, 15(3), 513-533.
- Frischkorn, G. T., Greiff, S., & Wüstenberg, S. (2014). The development of complex problem solving in adolescence: A latent growth curve analysis. *Journal of Educational Psychology*, 106(4), 1007-1020. doi:10.1037/a0037114
- Geiser, C. (2012). *Data analysis with Mplus*. Guilford Press.
- Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete non-normal data. *Structural Equation Modeling*, 10(1), 47-79.
- Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. H. (2009). A Latent Curve Model of parental motivational practices and developmental decline in math and science academic intrinsic motivation. *Journal of Educational Psychology*, 101(3), 729-739. doi:10.1037/a0015084
- Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications.

- Hallquist, M., & Wiley, J. (2014). MplusAutomation: Automating Mplus model estimation and interpretation. Retrieved from <https://cran.r-project.org/web/packages/MplusAutomation/MplusAutomation.pdf>
- Hancock, G. R., & Mueller, R. O. (Eds.). (2013). *Structural equation modeling: A second course*. LAP.
- Hohmann, L., Holtmann, J., & Eid, M. (2018). Skew t Mixture Latent State-Trait Analysis: A Monte Carlo Simulation Study on Statistical Performance. *Frontiers in Psychology*, 9, 1323. <http://doi.org/10.3389/fpsyg.2018.01323>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. Routledge.
- Hu, L. T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112(2), 351-362.
- Kaplan, D. (2008). *Structural equation modeling: Foundations and extensions* (Vol. 10). Sage Publications.
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Publications.
- Lai, M. H. C., & Zhang, J. (2017). Evaluating fit indices for multivariate t-based structural equation modeling with data contamination. *Frontiers in Psychology*, 8, 1-13. doi:10.3389/fpsyg.2017.01286
- Lee, S. X., Leemaqz, K. L., & McLachlan, G. J. (2016). A block EM algorithm for multivariate skew normal and skew t-mixture models. *arXiv preprint arXiv:1608.02797*.

- Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of non-normality in structural equation modeling. *Structural equation modeling*, 12(1), 1-27.
- Leite, W. L. (2007). A comparison of latent growth models for constructs measured by multiple items. *Structural Equation Modeling*, 14(4), 581-610.
- Li, J. & Lomax, R. G. (2017). Effects of missing data methods in SEM under conditions of incomplete and non-normal Data. *The Journal of Experimental Education*, 85(2), 31-258.
- Lin, T. I., & Lee, J. C. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Stat Med*, 27(9), 1490-1507.
doi:10.1002/sim.3026
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2), 130-149.
- Maronna, R.A., Martin, R.D. & Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons, Inc., NewYork,
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(3), 383-394.
- Mehta, P. D., & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological methods*, 5(1), 23-43.
- Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107-122.

- Miccéri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156–166.
- Mok, M. M. C., McInerney, D. M., Zhu, J., & Or, A. (2015). Growth trajectories of mathematics achievement: Longitudinal tracking of student academic progress. *British Journal of Educational Psychology*, 85(2), 154-171.
doi:10.1111/bjep.12060
- Montfort, K. v., Mooijaart, A., & Meijerink, F. (2009). Estimating structural equation models with non-normal variables by using transformations. *Statistica Neerlandica*, 63(2), 213-226. doi:10.1111/j.1467-9574.2009.00420.x
- Muthén, B. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. Collins, & J. Horn (Eds.), *Best Methods for the Analysis of Change. Recent Advances, Unanswered Questions, Future Directions* (pp. 1-17). Washington DC: American Psychological Association.
- Muthén, B. O. , & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2, 371–402.
- Muthén, B., & Asparouhov, T. (2015). Growth mixture modeling with non-normal distributions. *Statistics in medicine*, 34(6), 1041-1058.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431-462.
- Muthén, L. K., & Muthén, B. O. (2016). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1), 90-100.
- Petersen, J. L., & Hyde, J. S. (2017). Trajectories of self-perceived math ability, utility value and interest across middle school as predictors of high school math performance. *Educational Psychology*, 37(4), 438-456.
- Phan, H. P. (2012). The development of English and mathematics self-efficacy: A Latent Growth Curve Analysis. *Journal of Educational Research*, 105(3), 196-209.
doi:10.1080/00220671.2011.552132
- Pornprasertmanit, S., Miller P. and Schoemann, A. (2016). simsem: SIMulated Structural Equation Modeling. R package version 0.5-13. <https://CRAN.R-project.org/package=simsem>.
- Powell, D. A., & Schafer, W. D. (2001). The robustness of the likelihood ratio chi-square test for structural equation models: A meta-analysis. *Journal of Educational and Behavioral Statistics*, 26(1), 105-132.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C. & Briggs, N. E. (2008). *Quantitative Applications in the Social Sciences: Latent growth curve modeling* Thousand Oaks, CA: SAGE Publications Ltd doi: 10.4135/9781412984737
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. URL <http://www.jstatsoft.org/v48/i02/>.
- Satorra, A., & Bentler, P. M. (1988, August). Scaling corrections for chi-square statistics in covariance structure analysis. In *Proceedings of the American Statistical Association* (Vol. 1, pp. 308-313).

- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage Publications.
- Savalei, V. (2008). Is the ML chi-square ever robust to non-normality? A cautionary note with missing data. *Structural Equation Modeling*, 15(1), 1-22.
- Savalei, V., & Falk, C. F. (2014). Robust two-stage approach outperforms robust full information maximum likelihood with incomplete non-normal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 280-302.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147-177.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- Shin, T., Davidson, M. L., & Long, J. D. (2009). Effects of missing data methods in structural equation modeling with non-normal longitudinal data. *Structural Equation Modeling*, 16(1), 70-98.
- Tong, X., & Zhang, Z. (2012). Diagnostics of Robust Growth Curve Modeling Using Student's t Distribution. *Multivariate behavioral research*, 47(4), 493-518.
- Tong, X., & Zhang, Z. (2014). Abstract: Semiparametric Bayesian Modeling With Application in Growth Curve Analysis. *Multivariate Behav Res*, 49(3), 299.
doi:10.1080/00273171.2014.912928

- Tong, X., Zhang, Z. Y., & Yuan, K. H. (2014). Evaluation of Test Statistics for Robust Structural Equation Modeling with Non-normal Missing Data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 553-565.
doi:10.1080/10705511.2014.919820
- Ulitzsch, E., Holtmann, J., Schultze, M., & Eid, M. (2017). Comparing Multilevel and Classical Confirmatory Factor Analysis Parameterizations of Multirater Data: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(1), 80-103.
- User's Guide-assessments-PIAT Math.
<http://www.nlsinfo.org/site/childya/nlsdocs/guide/assessments/PIATMath.htm>
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48(3), 465-471.
- Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological bulletin*, 116(2), 363-381.
- You, S., & Sharkey, J. (2009). Testing a developmental–ecological model of student engagement: a multilevel latent growth curve analysis. *Educational Psychology*, 29(6), 659-684.
- Yuan, K. H., & Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 54(1), 161-175.

- Yuan, K. H., & Bentler, P. M. (2001). Effect of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, 54(1), 161-175.
- Yuan, K. H., & Chan, W. (2005). On nonequivalence of several procedures of structural equation modeling. *Psychometrika*, 70(4), 791.
- Yuan, K. H., & Zhang, Z. (2012). Robust structural equation modeling with missing data and auxiliary variables. *Psychometrika*, 77(4), 803-826.
- Yuan, K. H., & Zhong, X. (2013). Robustness of fit indices to outliers and leverage observations in structural equation modeling. *Psychological methods*, 18(2), 121-137.
- Yuan, K.-H., and Bentler, P. M. (1998). Structural equation modeling with robust covariances. *Sociological Methodology*, 28, 363–396. doi: 10.1111/0081-1750.00052
- Yuan, K.-H., and Zhang, Z. (2012). Structural equation modeling diagnostics using R package semdiag and EQS. *Structural Equation Modeling: A Multidisciplinary Journal*, 19, 683–702. doi: 10.1080/10705511.2012.713282
- Yuan, K.H., Bentler, P.M. and Chan, W. (2004). Structural equation modeling with heavy tailed distributions. *Psychometrika*. 69, 421–436.
- Zhang, Z. Y., Lai, K. K., Lu, Z. Q., & Tong, X. (2013). Bayesian Inference and Application of Robust Growth Curve Models Using Student's t Distribution. *Structural Equation Modeling-a Multidisciplinary Journal*, 20(1), 47-78. doi:10.1080/10705511.2013.742382

- Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior research methods*, 41(4), 1083-1094.
- Zhong, X., & Yuan, K. H. (2011). Bias and efficiency in structural equation modeling: Maximum likelihood versus robust methods. *Multivariate Behavioral Research*, 46(2), 229-265.
- Zu, J. & Yuan, K.H. (2010). Local influence and robust procedures for mediation analysis. *Multivariate Behavioral Research*, 45, 1–44.

APPENDIX

Table 1 Empirical Illustration - PIAT Descriptive Statistics

	N	Mean	SD	Skewness		Kurtosis	
Time1	1621	67.486	16.802	-0.617	0.061	3.614	0.121
Time2	1598	72.253	17.099	-0.797	0.061	3.768	0.122
Time3	1514	75.183	17.008	-0.882	0.063	4.054	0.126
Time4	1516	78.040	16.584	-0.887	0.063	3.866	0.126
Time5	1461	79.988	16.854	-1.063	0.064	3.449	0.128

Table 2 PIAT Tests of Normality

Kolmogorov-Smirnov				Shapiro-Wilk		
	Statistic	df	P value	Statistic	df	P value
Time1	0.044	1621	<0.001	0.974	1621	<0.001
Time2	0.080	1598	<0.001	0.952	1598	<0.001
Time3	0.110	1514	<0.001	0.939	1514	<0.001
Time4	0.114	1516	<0.001	0.929	1516	<0.001
Time5	0.147	1461	<0.001	0.901	1461	<0.001

Note. Significant results indicate that the distributions significantly differ from a normal curve.

Table 3 Covariance Coverage

	Time1	Time2	Time3	Time4	Time5
Time1	0.945				
Time2	0.881	0.931			
Time3	0.834	0.840	0.882		
Time4	0.836	0.830	0.802	0.883	
Time5	0.810	0.799	0.767	0.777	0.851

Note. Numbers represent proportion of data available for each combination of variables.

Table 4 Empirical Illustration Results - *st*MLE and *n*MLR Comparison

Parameter	<i>n</i> MLR		<i>st</i> MLE	
	Estimate	(SE)	Estimate	(SE)
α_1	68.457	(0.403)*	87.366	(0.597)*
α_2	2.893	(0.084)*	3.325	(0.126)*
φ_{11}	207.083	(9.633)*	44.083	(4.960)*
φ_{22}	2.646	(0.746)*	1.41	(0.237)*
φ_{21}	-1.901	(1.756)	-9.77	(0.955)*
θ_{11}	98.884	(11.053)*	48.303	(3.395)*
θ_{22}	75.321	(6.330)*	31.151	(2.255)*
θ_{33}	75.026	(8.097)*	29.109	(2.205)*
θ_{44}	62.941	(6.005)*	25.914	(2.022)*
θ_{55}	57.702	(7.775)*	17.261	(2.201)*
δ_1	-		-17.04	(0.524)*
δ_2	-		-0.782	(0.130)*
ν	-		3.131	(0.188)*
$X^2(10)$	55.989*		-	
<i>RMSEA</i>	0.052		-	
<i>CFI</i>	0.972		-	
<i>SRMR</i>	0.033		-	
<i>AIC</i>	59959.876		57730.965	
<i>BIC</i>	60014.354		57760.486	

* $p < .05$

Table 5 Simulation Conditions

Factor	Conditions
<i>Population Model</i>	Linear growth
	No growth
<i>Sample Size</i>	200
	500
	1000
<i>Distribution</i>	Normal
	Moderately non-normal
	Severely non-normal
<i>Missing Data</i>	Complete
	10% attrition
<i>Estimation</i>	<i>n</i> MLR
	<i>st</i> MLE

Note. There was a total of 72 conditions, or 36 conditions per estimation procedure.

Table 6 Convergence rates for stMLE across all Conditions

Condition			Sample Size		
			100	500	1000
Growth	<i>Normal</i>	Complete	98.8%	100.0%	100.0%
		10% Attrition	98.0%	99.8%	99.8%
	<i>Moderate</i>	Complete	92.0%	99.8%	99.8%
		10% Attrition	98.0%	100.0%	99.8%
	<i>Severe</i>	Complete	94.6%	100.0%	99.8%
		10% Attrition	91.6%	100.0%	99.8%
Zero Growth	<i>Normal</i>	Complete	98.2%	100.0%	100.0%
		10% Attrition	85.6%	100.0%	100.0%
	<i>Moderate</i>	Complete	91.6%	100.0%	100.0%
		10% Attrition	85.6%	100.0%	100.0%
	<i>Severe</i>	Complete	85.8%	100.0%	100.0%
		10% Attrition	85.8%	100.0%	100.0%

Note. Percentages are based on number of successful computations out of 500

samples for each condition.

Table 7 Intercept Variance Coverage Across all Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	95.0%	56.1%
	500	98.0%	49.2%
	1000	99.6%	16.4%
<i>Moderate</i>	100	94.6%	12.0%
	500	96.6%	0.0%
	1000	98.8%	0.0%
<i>Severe</i>	100	95.0%	12.9%
	500	97.8%	0.0%
	1000	99.6%	0.0%

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages.

Table 8 Slope Variance Coverage Across all Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	94.6%	73.3%
	500	96.4%	72.4%
	1000	97.4%	59.8%
<i>Moderate</i>	100	94.2%	68.90%
	500	96.4%	48.10%
	1000	96.4%	25.60%
<i>Severe</i>	100	95.2%	63.0%
	500	97.2%	47.7%
	1000	97.8%	32.0%

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages.

Table 9 Intercept/Slope Covariance Coverage Across all Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	94.0%	73.90%
	500	96.8%	81.00%
	1000	96.2%	63.20%
<i>Moderate</i>	100	53.8%	49.10%
	500	3.8%	18.00%
	1000	0.0%	1.40%
<i>Severe</i>	100	78.0%	42.10%
	500	43.6%	15.60%
	1000	13.8%	0.40%

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages.

Table 10 Mean Residual Coverage Across all Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	96.4%	83.3%
	500	98.3%	40.9%
	1000	98.6%	16.04%
<i>Moderate</i>	100	96.3%	79.2%
	500	98.2%	35.7%
	1000	99.1%	13.9%
<i>Severe</i>	100	96.4%	80.3%
	500	98.2%	37.5%
	1000	99.2%	14.8%

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages

Table 11 Intercept Mean Coverage Across all Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	95.0%	56.10%
	500	95.0%	83.00%
	1000	95.8%	90.80%
<i>Moderate</i>	100	94.8%	11.70%
	500	94.8%	0.00%
	1000	95.6%	0.00%
<i>Severe</i>	100	95.0%	15.20%
	500	95.0%	0.00%
	1000	96.0%	0.00%

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages

Table 12 Slope Mean Coverage Across all Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	93.4%	78.00%
	500	95.0%	70.90%
	1000	95.8%	52.20%
<i>Moderate</i>	100	92.6%	64.40%
	500	95.2%	84.20%
	1000	95.6%	89.60%
<i>Severe</i>	100	93.0%	75.70%
	500	95.0%	82.20%
	1000	95.8%	80.40%

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages

Table 13 Intercept/Slope Covariance Type I Error Rates Across all Zero Growth Conditions

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	0.050	0.259
	500	0.056	0.038
	1000	0.048	0.030
<i>Moderate</i>	100	0.054	0.607
	500	0.078	0.916
	1000	0.050	0.998
<i>Severe</i>	100	0.046	0.611
	500	0.068	0.886
	1000	0.056	0.597

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages.

Table 14 Slope Mean Type I Error Rates Across all Zero Growth Condition

Distribution	Sample Size	Estimation	
		<i>n</i> MLR	<i>st</i> MLE
<i>Normal</i>	100	0.076	0.367
	500	0.052	0.166
	1000	0.054	0.090
<i>Moderate</i>	100	0.084	0.366
	500	0.052	0.828
	1000	0.048	0.986
<i>Severe</i>	100	0.086	0.341
	500	0.052	0.814
	1000	0.050	0.317

Note. Results presented for complete data conditions. 10% attrition does not notably change the coverage percentages.

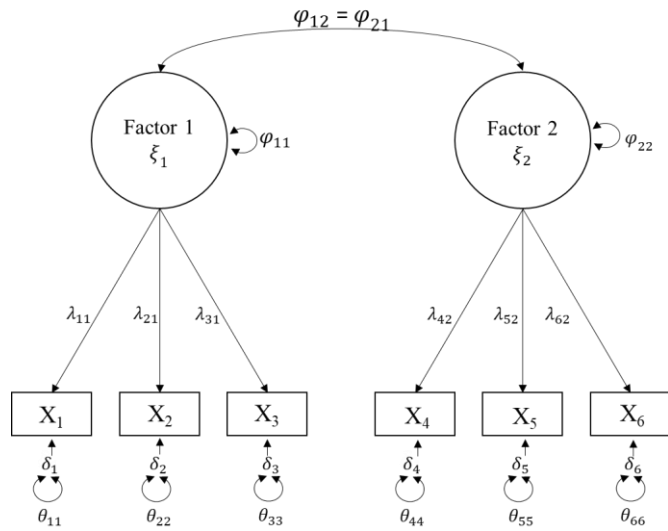


Figure 1 Confirmatory Factor Analysis (CFA) Model. A two-factor model with three indicators per factor and no cross loadings

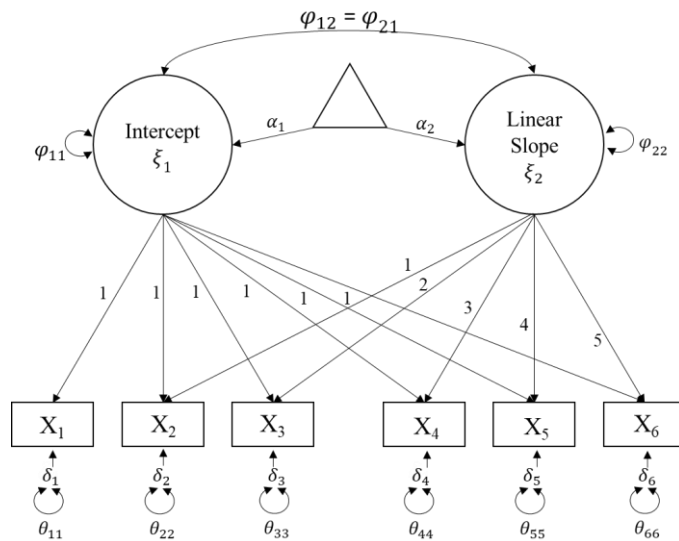


Figure 2 Latent Growth Curve Model (LGM). Linear growth and uncorrelated residuals and six measurement occasions

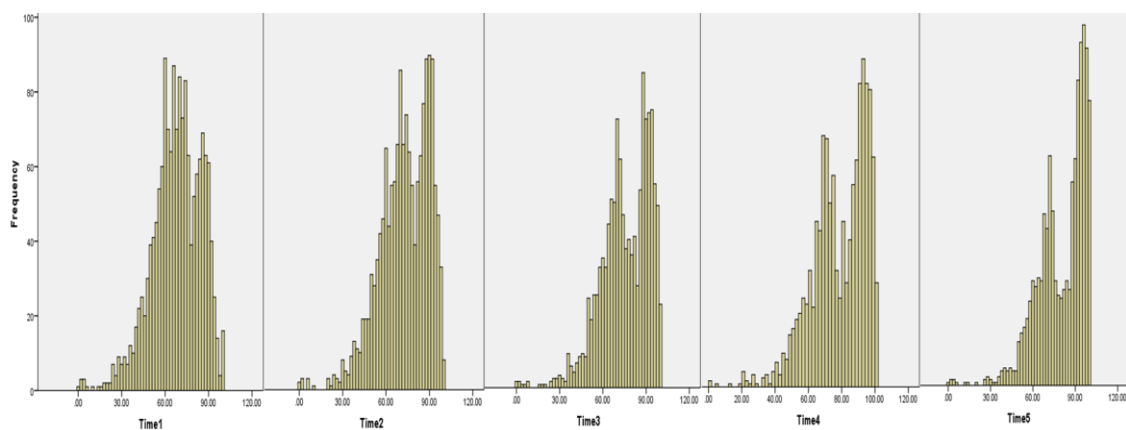


Figure 3 Histograms of observed PIAT scores at each time point

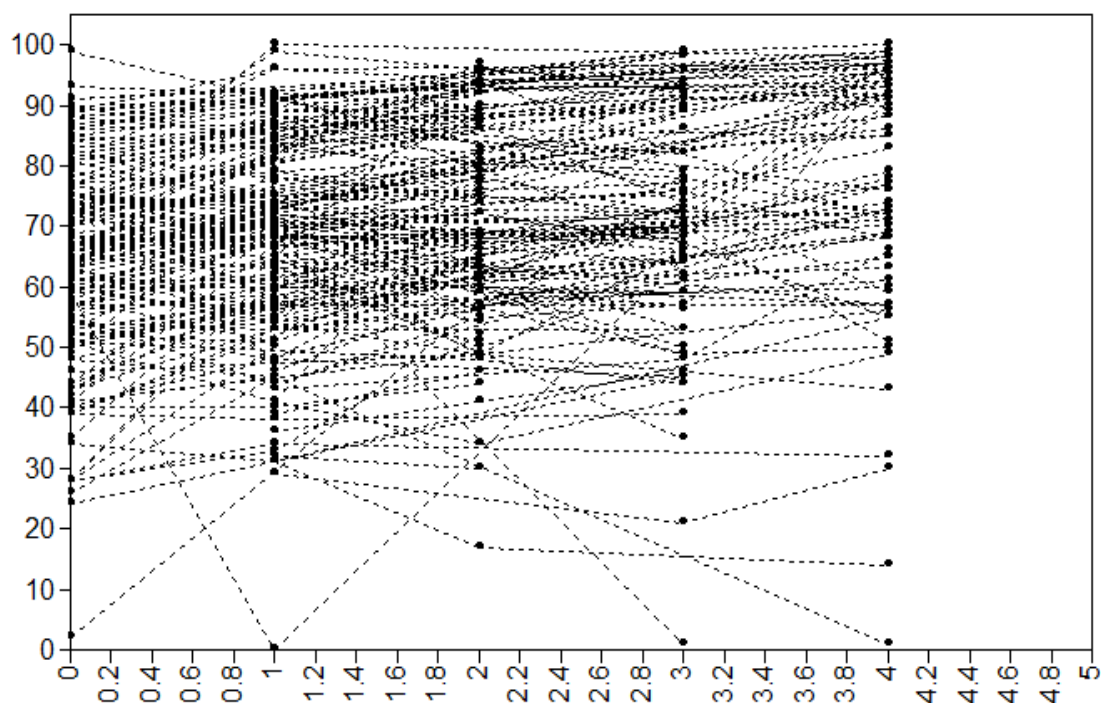


Figure 4 Individual observed growth plots of PIAT across five time points for 200 randomly selected individuals

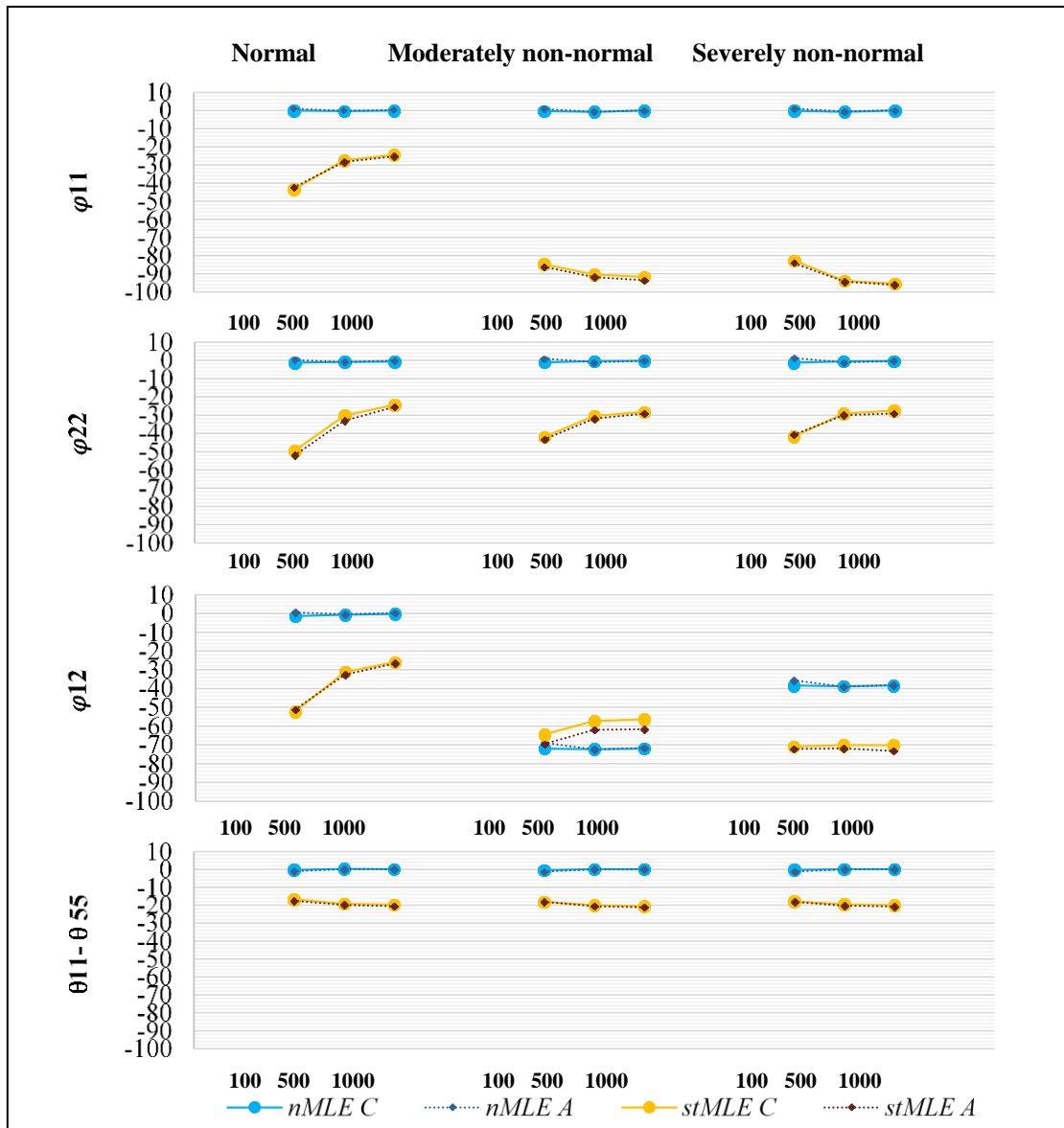


Figure 5 Random effects relative bias for growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents relative bias. Relative bias for residuals is averages across all five residuals per condition. C = complete data; A = 10% attrition.

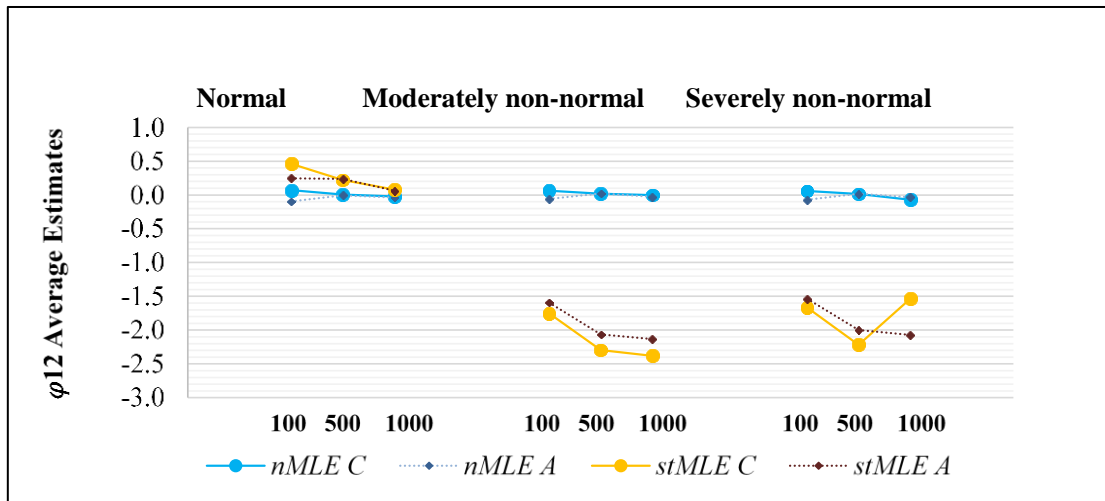


Figure 6 Growth factors covariance relative bias for zero growth conditions by estimation method. The X-axis represents sample size, and Y-axis represents relative bias. C = complete data; A = 10% attrition.

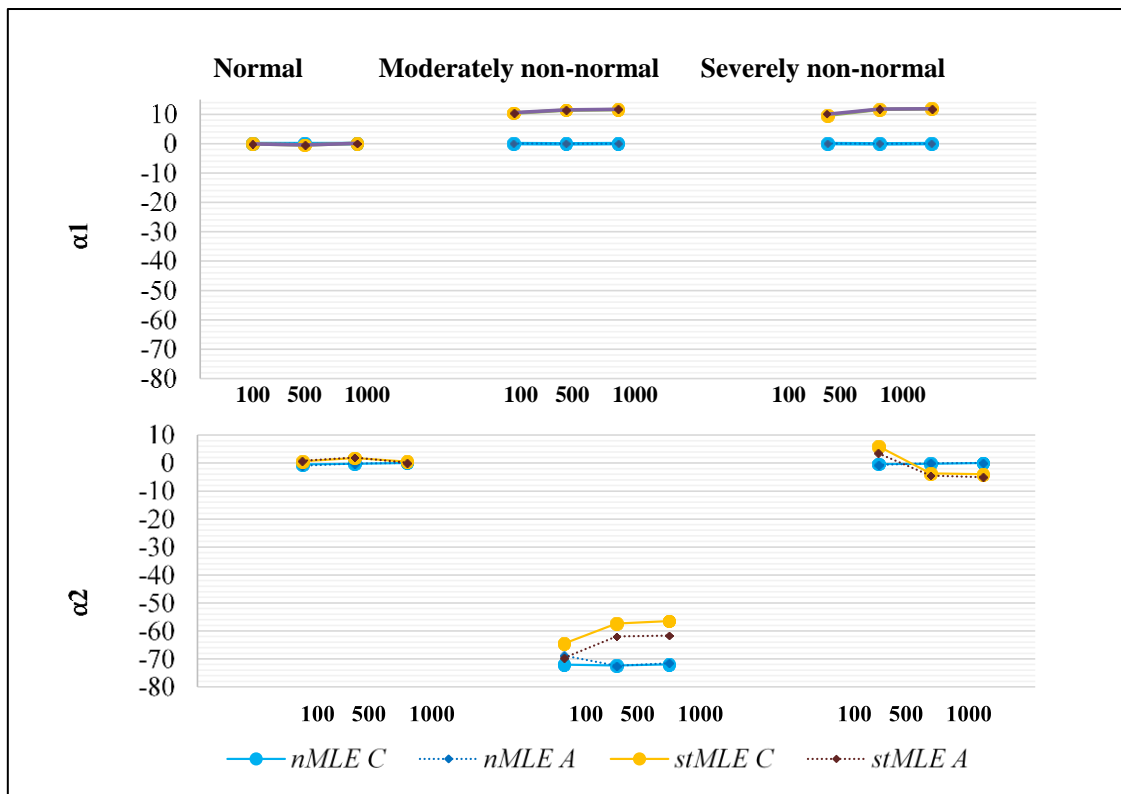


Figure 7 Fixed effects relative bias for growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents relative bias. C = complete data; A = 10% attrition.

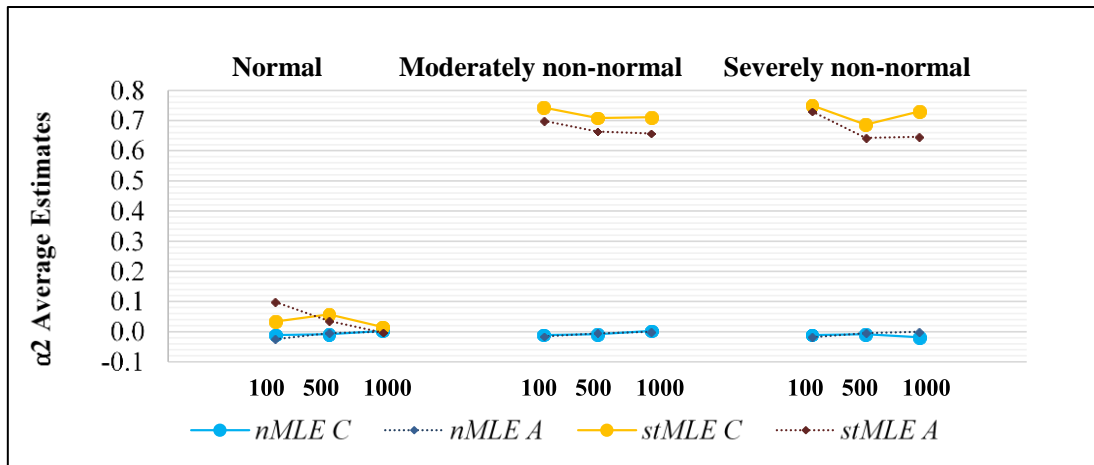


Figure 8 Slope mean relative bias for growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents relative bias. C = complete data; A = 10% attrition.

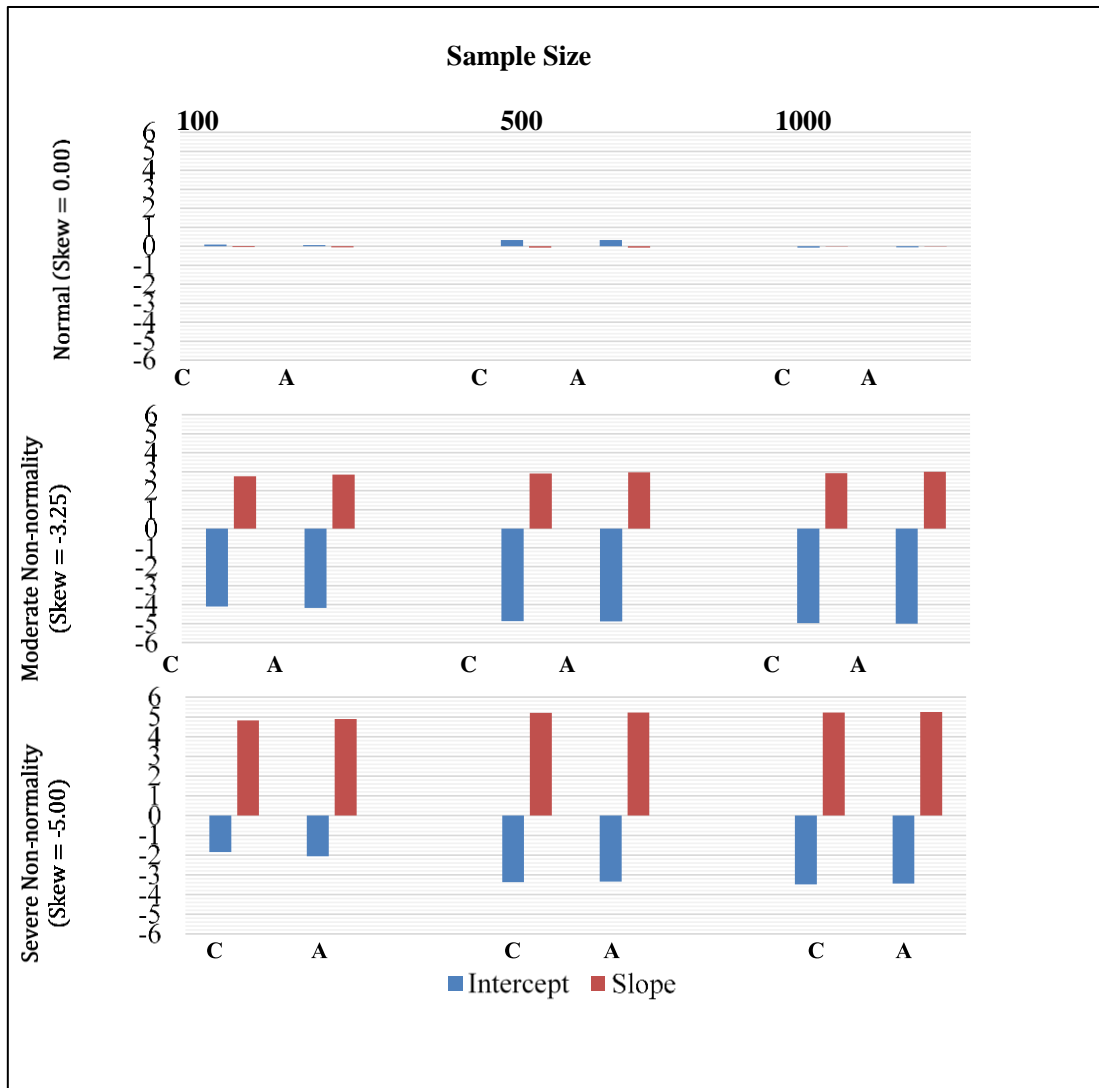


Figure 9 Difference between estimated skewness and population skewness for Growth Condition. Y-axis represents difference between estimated parameter and population value. X-axis represents sample size. C = complete data; A = 10% attrition.

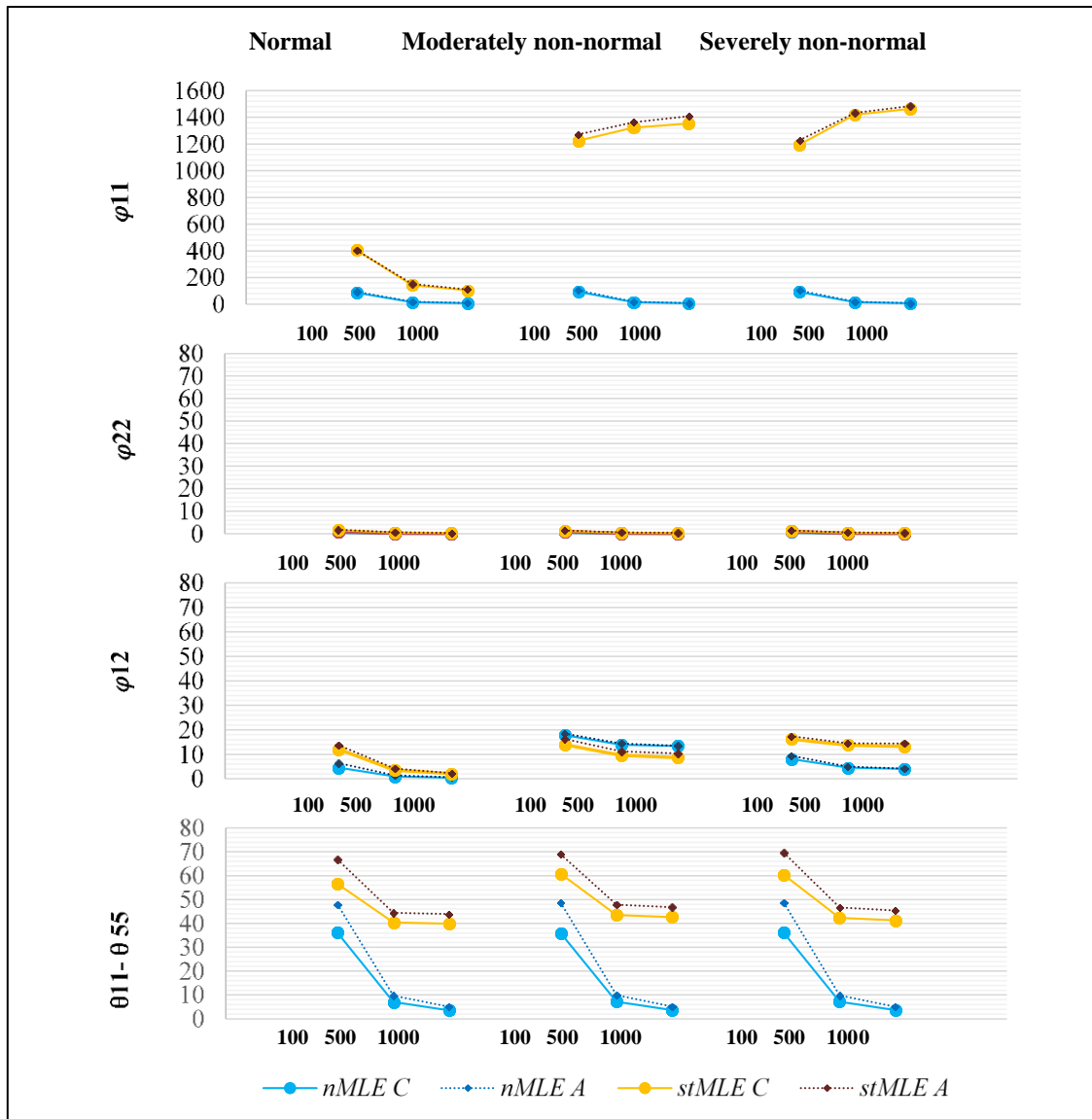


Figure 10 Random effects MSE for growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents MSE. MSE for residuals is averages across all five residuals per condition. C = complete data; A = 10% attrition.

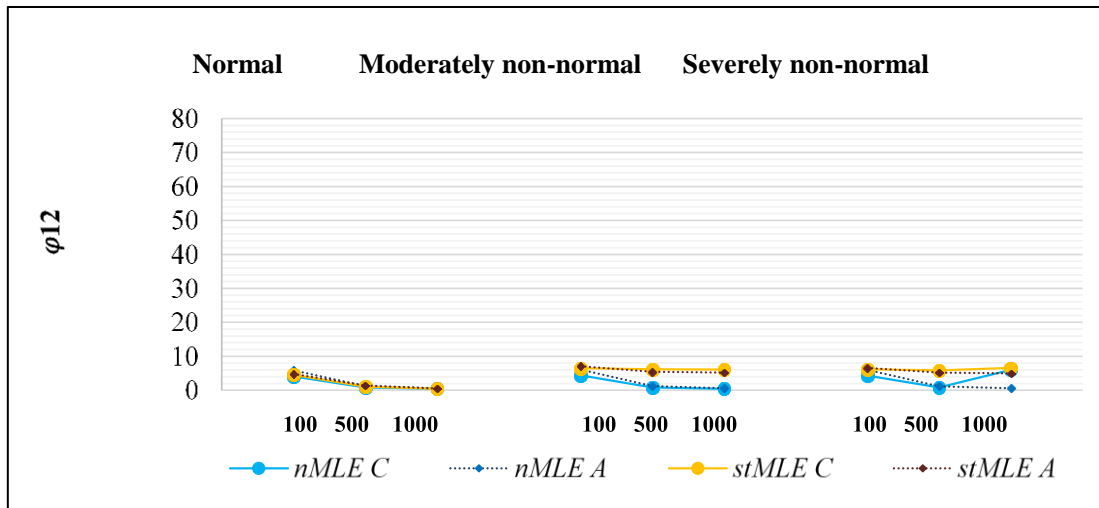


Figure 11 Growth factors covariance MSE for zero growth conditions by estimation method. The X-axis represents sample size, and Y-axis represents MSE. C = complete data; A = 10% attrition.

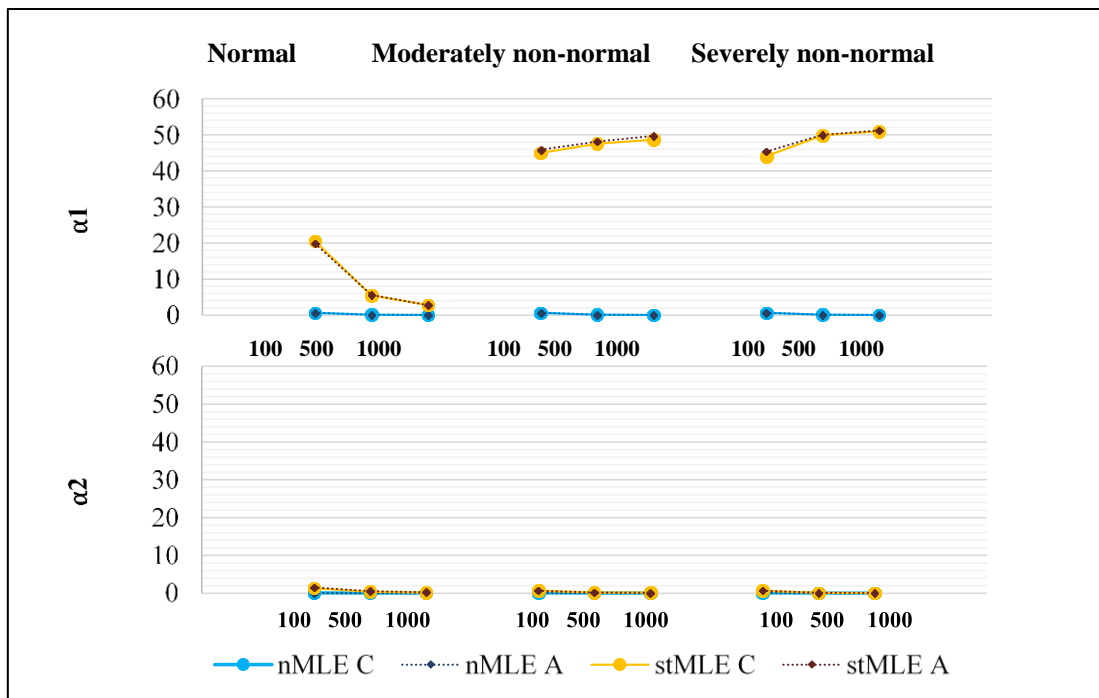


Figure 12 Fixed effects MSE for growth conditions by estimation method

X-axis represents sample size, and Y-axis represents MSE. C = complete data; A = 10% attrition.

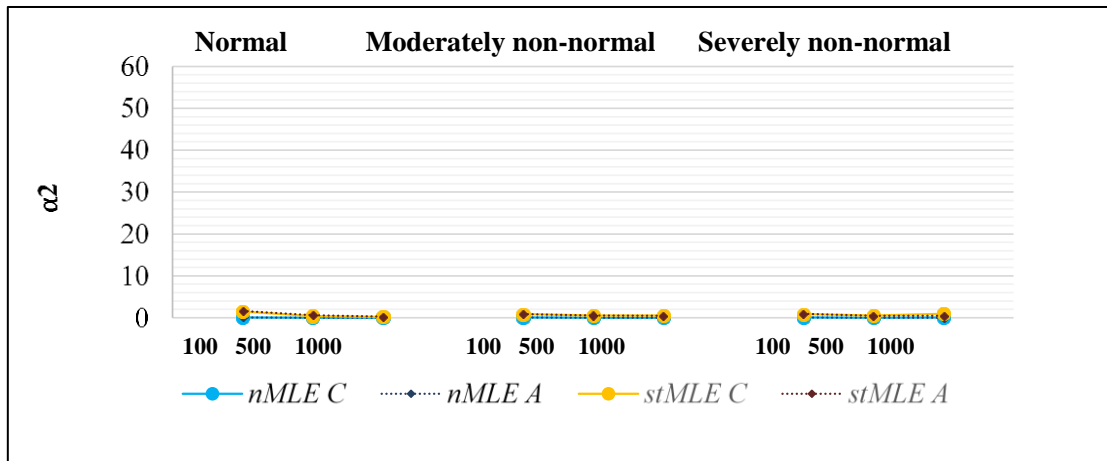


Figure 13 Slope mean MSE for zero growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents MSE. C = complete data;

A = 10% attrition.

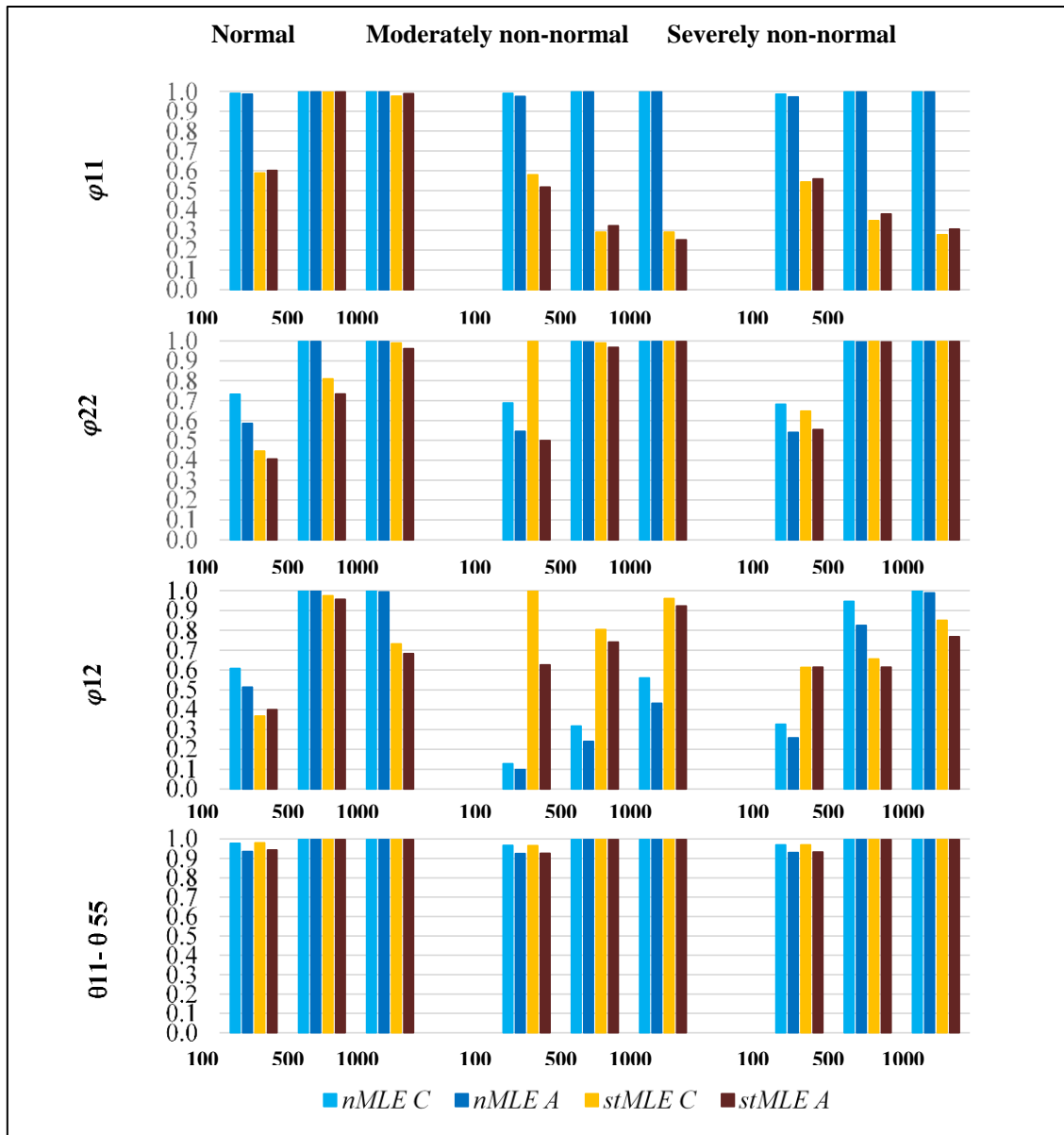


Figure 14 Random effects power rates for growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents power. Power for residuals is averages across all five residuals per condition. C = complete data; A = 10% attrition.

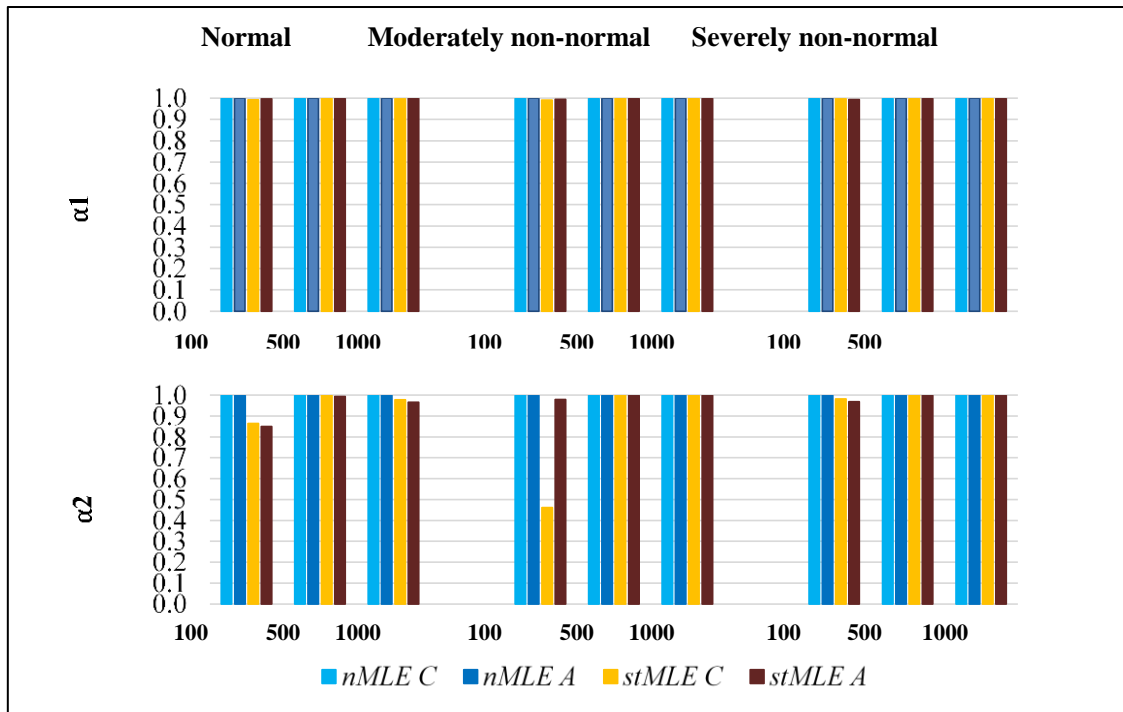


Figure 15 Fixed effects power for growth conditions by estimation method

The X-axis represents sample size, and Y-axis represents power. C = complete data;
A = 10% attrition.